



Previsão de valores de humidade do solo com dados de rede de sensores sem fios (wireless) em área aberta e previsão meteorológica da internet para aplicação na agricultura

por

André Manuel Marques Ferreira Lino

Dissertação para obtenção do grau de Mestre em Análise de Dados e Sistemas de Apoio à Decisão pela Faculdade de Economia do Porto

Orientada por: Professor João Gama
Co-orientada por: Professor João Moreira

setembro de 2017

Nota Biográfica

André Manuel Marques Ferreira Lino nasceu a 2 de março de 1986, na cidade de Santa Maria da Feira, onde ainda reside atualmente.

Estudou em Santa Maria da Feira até 2004 quando terminou o ensino secundário em Económico-social, na Escola Secundária de Santa Maria da Feira, e ingressou no ensino universitário nesse mesmo ano, concluindo a sua licenciatura em Gestão pela Faculdade de Economia da Universidade do Porto em 2008.

Durante o percurso académico e profissional foi procurando adquirir conhecimentos na área de eletrónica, automatização e programação pois sempre teve o “bichinho” da inovação e uma preocupação pela eficiência aplicação dos recursos. Foi por essa curiosidade constante que decidiu ingressar no mestrado em Análise de Dados e Sistemas de Apoio à Decisão pela Faculdade de Economia do Porto em 2014.

Todo o seu percurso profissional tem sido desenvolvido no departamento financeiro, relacionado com faturação e inventários.

Agora procura aplicar os conhecimentos do mestrado na sua via profissional e ao mesmo tempo aliar esses conhecimentos a outra área de interesse que sempre demonstrou ter, a agricultura.

Agradecimentos

Ao Professor Doutor João Gama, deixo dois agradecimentos, um pelo mestrado que dirige, pela boa energia que faz passar para os alunos e nunca desistir de ninguém, outro agradecimento mais pessoal pelo trabalho desenvolvido comigo, toda a orientação e persistência.

Ao Professor Doutor João Moreira por ter aceitado o meu tema e pela primeira orientação que me fez acreditar ser possível.

Às entidades patronais que ajudaram na conciliação dos estudos com a vida profissional facilitando na exigência e nas horas de trabalho.

À minha família que me acompanhou desde o início, em especial à minha esposa por todo o esforço feito para que juntos conseguíssemos trabalhar, estudar, tratar da nossa casinha e o melhor de tudo fazer a nossa família crescer.

Ao meu filho mais velho, um agradecimento especial, por todo o tempo que devíamos ter passado juntos e que lhe fiquei a dever por ter de me ausentar para as aulas e chegar a casa quando já estava a dormir.

Por último, quero agradecer ao meu pai, que mesmo não sendo possível acompanhar esta etapa da minha vida, sem ele muita coisa não seria possível. Obrigado pai.

Resumo

A evolução humana vem da necessidade da procura de ferramentas que simplifiquem a vida e que tornem as tarefas mais eficientes, menos dispendiosas monetariamente ou mais “leves” para o ser humano. Aliada a essa eterna procura tem existido ultimamente o desenvolvimento da inteligência artificial, aumento da capacidade de conectividade e processamento de dados.

Hoje é possível ter outras armas para que a vida humana não seja tão destrutiva para o nosso planeta e para vivermos melhor com a escassez e adversidades que a natureza nos faz enfrentar. Os recursos são cada vez mais escassos e as necessidades são infinitas e necessitam de maiores quantidades de recursos para o presente dado o crescimento da população humana.

Este trabalho tem como objetivo proporcionar uma ferramenta para apoio na decisão relativa à rega de plantas através de um modelo preditivo da humidade do solo. Foi com esse objetivo que foi implementada uma rede sem fios entre emissor e estação base para recolha de dados. O emissor é o nó da rede onde estão instalados os sensores de humidade e temperatura. O outro nó é a estação base que recebe as leituras dos sensores e agrega a informação fornecida pela *Application Programming Interface(API)* de *weather forecast* utilizada formando assim um registo para a base de dados para aplicação do modelo preditivo.

Além de criar um modelo capaz de prever o comportamento da humidade da terra, o trabalho realizado avalia se a conjugação de dados de sensores no terreno com a previsão meteorológica disponível da internet torna o modelo preditivo melhor e se é possível reduzir o número de sensores instalados no nó emissor.

Para criar o modelo preditivo foram testados e avaliados os algoritmos *Multilayer Perceptron(MLP)*, *Regression Trees* e *Polynomial Regression*.

Abstract

Neste trabalho é abordada a importância das redes de sensores, o aumento da procura e da sua aplicação. É explicada a diferença entre as redes com fios das redes sem fios e o que condiciona a aplicação de técnicas de *data mining*. Na revisão de literatura foram analisadas técnicas de previsão de dados meteorológicos. Na Metodologia de investigação é apresentada a base de dados, criada por dados extraídos de rede de sensores wireless (com dados meteorológicos – temperatura e humidade quer do ar quer do solo) no terreno e de uma aplicação de *forecast* de previsões meteorológicas. São igualmente explicadas as técnicas aplicadas na base de dados. No capítulo de análise de dados são aplicados os modelos de previsão e posteriormente avaliados. É determinado qual o que tem melhor desempenho na previsão da humidade do solo nas próximas 4 horas. No final são apresentados alguns temas para aprofundar o estudo bem como algumas das aplicações do modelo.

Keywords: Wireless Sensor Networks, Previsão, Data Mining, Machine learning, weather forecast. Temperatura, Humidade.

Índice

| | |
|---|-----------|
| Nota Biográfica | 2 |
| Agradecimentos..... | 3 |
| Resumo | 4 |
| Abstract | 5 |
| Índice | 6 |
| Siglas | 11 |
| I. Introdução..... | 12 |
| Enquadramento..... | 12 |
| Traços gerais da metodologia de investigação | 13 |
| Estrutura da Dissertação | 15 |
| II. Revisão Bibliográfica | 16 |
| Velocidade Vento | 16 |
| Temperatura interior..... | 16 |
| Temperatura exterior | 17 |
| Humidade no solo..... | 18 |
| Valores meteorológicos..... | 19 |
| ANN vs Modelos Estatísticos | 20 |
| SVM Vs NN..... | 20 |
| III. Metodologia de Investigação | 22 |
| Rede de Sensores Wireless..... | 22 |
| Recolha de dados e criação automática da base de dados..... | 27 |
| Tratamento da base de dados | 29 |
| <i>Dados de sensores no terreno</i> | <i>30</i> |
| <i>Dados recebidos via API</i> | <i>31</i> |

| | |
|--|----|
| <i>Dados acrescentados</i> | 33 |
| IV. Análise de Dados | 37 |
| Multilayer perceptron/ Perceptron Multicamada | 38 |
| Regression trees / Árvores de Regressão | 39 |
| Polynomial Regression / Regressão Polinomial..... | 40 |
| Resultados | 41 |
| <i>SH1 - primeira previsão horária da humidade do solo</i> | 41 |
| <i>SH2 - segunda previsão horária da humidade do solo</i> | 42 |
| <i>SH3 - terceira previsão horária da humidade do solo</i> | 43 |
| <i>SH4 - quarta previsão horária da humidade do solo</i> | 44 |
| <i>Outros estudos</i> | 46 |
| V. Conclusões e Pesquisas Futuras | 52 |
| Referências Bibliográfica | 54 |
| Anexos..... | 56 |

Índice de Figuras

| | |
|--|----|
| Figura 1 - Sensor Network | 12 |
| Figura 2 – OUTPUT 4 previsões horárias da humidade..... | 14 |
| Figura 3 – Workflow da Rede | 23 |
| Figura 4 - Composição do Nó | 24 |
| Figura 5 - Nó Base - Recetor e PC | 27 |
| Figura 6 - Dados Recebidos no Recetor e comunicados via serial..... | 27 |
| Figura 7 - Dados da API no ficheiro .JSON | 28 |
| Figura 8 - CSV criado com as leituras..... | 29 |
| Figura 9 - MetaNode para retirar as grandes variações negativas | 36 |
| Figura 10 - WORKFLOW do KNIME para aplicação de Modelos para Variável SH4 | 37 |
| Figura 11 - Configuração do Nó de aprendizagem MLP no KNIME | 38 |
| Figura 12 - Valores do algoritmo MLP para a variável SH4..... | 38 |
| Figura 13 - Gráfico KNIME da variável SH4 e da Previsão SH4 do modelo Regression Trees | 39 |
| Figura 14 - Desenho do nó da árvore para a variável SH4..... | 39 |
| Figura 15 - Configuração do Nó de aprendizagem POLYNOMIAL REGRESSION no KNIME..... | 40 |
| Figura 16 - Variáveis da nova abordagem..... | 50 |

Índice de Tabelas

| | |
|--|----|
| Tabela 1 - Diferença entre Rede com fios e Rede sem fios..... | 26 |
| Tabela 2 – Variáveis da base de dados | 30 |
| Tabela 3 - Estatísticas de Variáveis de Sensores..... | 31 |
| Tabela 4 - Erros encontrados na comunicação | 34 |
| Tabela 5 - Leituras em falta..... | 34 |
| Tabela 6 - Leituras em falta - detalhe julho..... | 35 |
| Tabela 7 - DATA SET..... | 36 |
| Tabela 8 - Valores dos coeficientes do algoritmo POLYNOMIAL REGRESSION para a variável SH4 | 40 |
| Tabela 9 - SH1 no TREES..... | 41 |
| Tabela 10 - SH1 no POLYNO..... | 41 |
| Tabela 11 - SH1 no MLP..... | 41 |
| Tabela 12 - SH1 sem API no MLP..... | 41 |
| Tabela 13 - SH1 sem API no TREES..... | 41 |
| Tabela 14 - SH1 sem API no POLYNO..... | 41 |
| Tabela 15 - SH2 no TREES..... | 42 |

| | |
|--|----|
| Tabela 16 - SH2 no POLYN. | 42 |
| Tabela 17 - SH2 no MLP..... | 42 |
| Tabela 18 - SH2 sem API no MLP..... | 42 |
| Tabela 19 - SH2 sem API no POLYN..... | 42 |
| Tabela 20 - SH2 sem API no TREES..... | 42 |
| Tabela 21 - SH3 no TREES..... | 43 |
| Tabela 22 - SH3 no POLYNO..... | 43 |
| Tabela 23 - SH3 no MLP..... | 43 |
| Tabela 24 - SH3 sem API no POLYNO..... | 43 |
| Tabela 25 - SH3 sem API no TREES..... | 43 |
| Tabela 26 - SH3 sem API no MLP..... | 43 |
| Tabela 27 - SH4 no MLP..... | 44 |
| Tabela 28 - SH4 no TREES..... | 44 |
| Tabela 29 - SH4 no POLYNO..... | 44 |
| Tabela 30 - SH4 sem API no POLYNO..... | 44 |
| Tabela 31 - SH4 sem API no TREES..... | 44 |
| Tabela 32 - SH4 sem API no MLP..... | 44 |
| Tabela 34 - SH4 no MLP sem grandes variações..... | 47 |
| Tabela 33 - SH4 no POLYNOMIAL sem grandes variações..... | 47 |
| Tabela 36 - SH4 no MLP aumento data treino..... | 47 |
| Tabela 35 - SH4 no POLYNOMIAL aumento data treino..... | 47 |
| Tabela 37 - SH4 no MLP aproximação data treino..... | 48 |
| Tabela 38 - SH4 no POLYNOMIAL aproximação data treino..... | 48 |
| Tabela 41 - SH4 no POLYNOMIAL sem "Airhum", "Airtemp" e "Soiltemp" | 49 |
| Tabela 42 - SH4 no MLP sem "Airhum", "Airtemp" e "Soiltemp" | 49 |

Índice de Gráficos

| | |
|--|----|
| Gráfico 1 – Quebra dos valores observados com acontecimento marcante SH1 para SH2 (dif_2) | 36 |
| Gráfico 2 - Comparação de resultados do MLP | 45 |
| Gráfico 3 - Comparação de resultados do TREES | 45 |
| Gráfico 4 - Comparação de resultados do POLYNOMIAL | 46 |
| Gráfico 5 - Resumo nova abordagem MLP com dados API | 50 |
| Gráfico 6 - Resumo nova abordagem TREES com dados API | 50 |
| Gráfico 7 - Resumo nova abordagem POLYNOMIAL com dados API | 50 |
| Gráfico 8 - Resumo nova abordagem MLP sem dados API..... | 51 |
| Gráfico 9 - Resumo nova abordagem TREES sem dados API..... | 51 |
| Gráfico 10 - Resumo nova abordagem POLYNOMIAL sem dados API..... | 51 |

Índice de Anexos

| | |
|---|----|
| Anexo 1 - Variáveis retiradas | 56 |
| Anexo 2 - Exemplo de retirar possíveis erros/outliers | 56 |
| Anexo 3 - Microcontrolador utilizado - Arduino Uno Rev3 | 57 |
| Anexo 4 - características do microcontrolador..... | 57 |
| Anexo 5 - sensor de humidade do solo - G305..... | 57 |
| Anexo 6 - Sensor temperatura solo - DS18B20 | 58 |
| Anexo 7 - Sensor humidade e temperatura do ar - DHT22 (AM2303) | 58 |
| Anexo 8 - Modulo de Comunicação - NRF24L01 + | 59 |
| Anexo 9 – CÓDIGO para POWERSHELL..... | 60 |
| Anexo 10 - Nó emissor aplicado no Vaso | 60 |
| Anexo 11 - Nó emissor..... | 61 |
| Anexo 12 - Nó emissor - detalhe do sensor da humidade e temperatura do ar | 62 |
| Anexo 13 - detalhe do sensor da humidade e temperatura do solo | 62 |
| Anexo 14 - PC esperando por leitura de dados do recetor | 63 |
| Anexo 15 - PC após receber leitura do recetor - mostra dados recebidos | 63 |

Siglas

API – *Application Programming Interface* - Interface de Programação de Aplicações ou Interface de Programação de Aplicação

CSV - *Comma-Separated Values* – Valores Separado por Vírgula

FEP – Faculdade de Economia do Porto

IOT - *Internet Of Things* – Internet Das Coisas

JSON – *JavaScript Object Notation*

MAE – *Mean Absolute Error* – Média de Erros Absolutos

MLP – *MultiLinear Perceptron* – Perceptron Multicamadas

MSE – *Mean Squared Error* – Média de Erros Quadrados

SH – *Soil Humidity* – Humidade do Solo

SMS – *Short Message Service* – Serviço de Mensagens Curtas

SVM - *Support Vector Machine* – Máquina de Vetores de Suporte

URL – *Uniform Resource Locator* - Localizador Uniforme de Recursos

I. Introdução

Enquadramento

O mundo avança com a tecnologia e novas descobertas. Nas últimas décadas o chip e os processadores têm observado um desenvolvimento exponencial, em tamanho e capacidade de processamento. Proporcionando nos últimos anos o desenvolvimento e a aplicação de eletrónica em quase tudo que esta presente na nossa vida. A utilização de tablets e de comunicação via internet móvel tem batido records, os carros estão cada vez mais eletrónicos e menos mecânicos e os smartphones tiveram um crescimento exponencial deixando obsoletos os antigos telemóveis. Cada vez mais as operadoras de telecomunicação deixam de transmitir som e SMS e tem que responder ao crescimento dos dados.

O desenvolvimento da forma e velocidade de comunicação tem alterado o conceito de distância e torna a informação útil pois é possível receber em tempo real e tomar decisões. Observamos o aparecimento do conceito de *Internet Of Things – IoT* que consiste em “ligar” à rede web tudo (a casa, carro, sapatos, relógio, animal de estimação,...).

Começam a surgir as *smart cities* que controlam e monitorizam atividades em toda a cidade [25]. Todos estes sensores geram dados que sendo armazenados e tratados podem determinar padrões e comportamentos.

Neste estudo, o objetivo é conjugar estes avanços tecnológicos e prever o comportamento da humidade do solo. Conforme é visível na Figura 1 o solo foi monitorizado recorrendo a sensores, essa informação foi comunicada via wireless para uma estação base que agrega os dados recebidos aos dados fornecidos por uma API, completando assim o registo da leitura que fica numa base de dados utilizada para aprender o modelo preditivo.

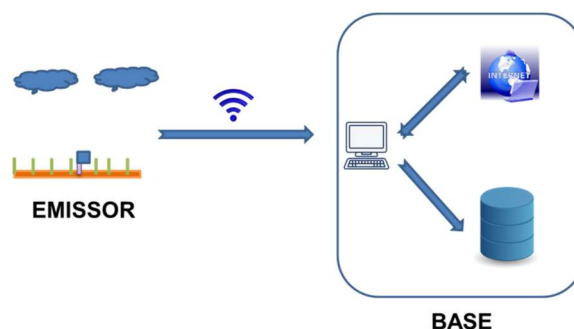


Figura 1 - Sensor Network

O *output* do modelo preditivo são quatro previsões horárias. Quando temos uma leitura dos sensores queremos saber qual será a humidade do solo daqui a uma hora, daqui a duas, três e quatro.

Traços gerais da metodologia de investigação

Após implementação da rede em espaço aberto (apenas com os dois nós) foi necessário armazenar e registar dados suficientes para treino e teste.

Durante a recolha dos dados foi necessário manter a terra em estudo com um grau de hidratação capaz de ser lida pelo sensor de humidade, por isso foi colocado o emissor num vaso grande onde foram colocadas beterrabas e foi necessário regar e manter a humidade necessária para o seu desenvolvimento. De salientar que a recolha de dados foi efetuada de 30 em 30 minutos durante os meses de abril a agosto. Foram meses que necessitaram de regadio, o país apresentou nesses meses seca extrema ou severa, ou seja, a precipitação foi muito rara. Foi acrescentada água ao vaso sem nenhuma regra definida, foi um comportamento aleatório para serem registados dados da terra em diferentes casos e para isso compreender melhor o comportamento da humidade na terra estudada.

Após obter os dados foi necessário **tratar** esses dados (uniformizando as unidades de medida e colocando principalmente como separador decimal o ponto”), para leitura no KNIME. Esta questão pode ser inserida no código de criação de base de dados, mas neste caso foi feito manualmente durante a análise dos dados. Posteriormente foram verificados os **dados nulos/erros** (dado ser uma rede *wireless*, pode existir muita interferência/ruído nas comunicações ou mesmo falhas da receção, o que provoca muitas inconsistências e dificuldade em criar o modelo) no caso em estudo ao monitorizar no final de abril a consistência dos dados recebidos na estação base, foi possível concluir que a base de dados apresentava muitas falhas. Com esta quantidade de falhas não era possível criar um método de previsão e posterior avaliação pois não existem leituras consecutivas para termos o registo da humidade do solo no momento 0, no momento +1 hora, + 2 horas, + 3 horas, +4 horas. Para resolver a ausência de registos foi necessário alterar a comunicação rádio estabelecida entre os nós da rede. Foi colocado no código da programação da estação base uma instrução para efetuar um *reset* após receber uma leitura ou executar *reset* após 50 minutos da última leitura recebida, garantido assim que a comunicação será estabelecida novamente na próxima escuta.

Relativamente a **dados inconsistentes**, parte dos dados têm origem em sensores ao “ar

livre”, por isso os dados podem conter variações elevadas sem explicação. Foi importante introduzir algum método de procura de *outliers* e fazer uma análise estatística para verificar a distribuição da variável.

Depois do tratamento dos dados e como que o objetivo é estudar o comportamento da terra face ao índice de humidade foi necessário acrescentar à base de dados a variável que queremos prever. Na figura 2 é possível compreender por gravura o *output* do estudo, e a necessidade de acrescentar as colunas com a variável que queremos prever SH1, SH2, SH3, SH4 .

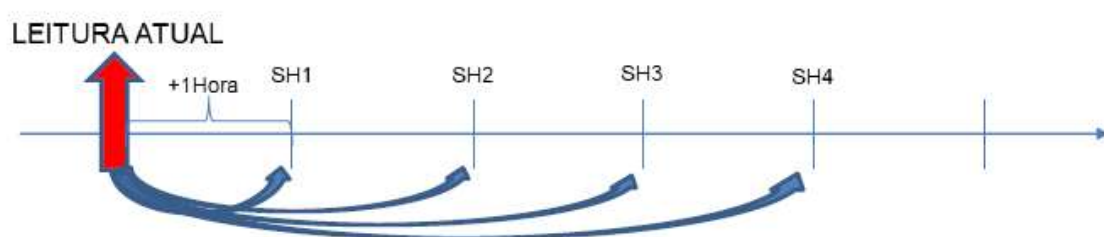


Figura 2 – *OUTPUT 4* previsões horárias da humidade

Foram também acrescentadas colunas onde foram calculadas as variações da humidade para cada leitura (feita a subtração da leitura N com a N+1) para determinar em que tempo aconteceu acontecimentos significantes em termos de aumento da humidade (rega/chuva).

Posteriormente foram aplicados vários modelos entre os quais: *MultiLayerPerceptron*, *Regression Trees*, *Polynomial Regression*. Na avaliação dos modelos foram utilizados os indicadores MSE e MAE. Esta última fase foi efetuada no sistema KNIME.

Objetivamente queremos dar resposta a qual o melhor modelo preditivo e se a informação que acrescentamos da previsão meteorológica tem peso significativo para a melhoria do modelo.

Estrutura da Dissertação

A dissertação inicia com a **introdução** onde é apresentada a razão do interesse do tema, a sua importância na atualidade e os traços gerais do estudo realizado.

No capítulo seguinte é explorada a **revisão de literatura**, foram lidos e explicados alguns estudos já realizados na previsão de dados meteorológicos. Foram verificados os modelos aplicados e quais os constrangimentos que enfrentaram.

No terceiro capítulo é apresentada a **metodologia de investigação**. São apresentadas as variáveis e como foram tratados os dados até serem incluídos no sistema KNIME para aplicação de modelos de previsão. Foram explicadas as técnicas para inconsistências, erros, nulos e explicadas as variáveis calculadas que foram acrescentadas.

No quarto capítulo é feita a **análise de dados**, são apresentados os resultados dos modelos aplicados e a sua comparação para determinar qual apresenta melhor avaliação.

O quinto, e último capítulo, englobam as **conclusões finais** relativas ao tema em estudo e **possíveis pesquisas futuras**.

II. Revisão Bibliográfica

A revisão bibliográfica foi centrada na necessidade de conhecer o trabalho efetuado na previsão de dados meteorológicos. Conhecer que modelos foram utilizados, as dificuldades na previsão e como estes são influenciados por outros fatores e quais fatores. Não foram abordados nenhuns estudos específicos relativo à criação de redes sem fios.

Velocidade Vento

Aplicação na previsão da velocidade do vento [15] em que o objetivo é prever o ganho de energia, a energia eólica é uma fonte sustentável de energia elétrica e que está em expansão.

Energia eólica é produzida ao ar livre e é proporcional ao cubo da velocidade do vento. E por isso a previsão da energia eólica é diretamente dependente da precisão da previsão da velocidade do vento.

Os modelos de previsão podem ser compostos em dois grupos: métodos de previsão com base física, em que são recolhidos os valores no terreno (vento, humidade, temperatura, pressão, etc), e são aplicados comportamentos físicos já estudados, e uma previsão com base estatística que estabelecem a relação entre os valores históricos de velocidade do vento, bem como os valores históricos e de previsão meteorológica.

Neste estudo é introduzido um novo procedimento, primeiro é aplicado *Clustering*, (*k-means*) selecionada em que os valores das variáveis são classificados em várias categorias com diferentes rótulos. Depois é feita uma regra de associação no qual é aplicado o algoritmo *Apriori*. São estabelecidos os padrões de relações entre as variáveis. De seguida é feita a previsão da velocidade do vento tendo como base a técnica “*Chaotic Time Series*”, e depois corrige os valores previstos utilizando as regras associadas descobertas anteriormente.

Temperatura interior

Este tema é apresentado no [16] e [17] está ligado com o objetivo de controlar a humidade interior controlando a temperatura interior de edifícios pois pode ser um problema grave se a humidade se aproximar de valores extremos, provocando elevados prejuízos degradando a construção e a eficiência energética. Além dos danos na estrutura o mais importante é obter um ar de qualidade e para isso é necessário evitar o excesso de humidade que criam fungos e consequentemente problemas de saúde.

Os métodos utilizados recorreram a *neural networks* para criar o método de previsão e ação, sendo que o trabalho apresentado [16] complementa o *neural networks* com o *genetic Algorithm*,

Existem modelos físicos e bastante complexos para estudar a relação temperatura-humidade mas consome bastante tempo e recursos computacionais. As *neural networks* conseguem em minutos obter valores para um edifício médio. Foram utilizadas *neural networks*/redes neurais com três camadas, a escolha do número de camadas é a primeira dificuldade e tem grande impacto sobre a implementação, treinar o algoritmo e avaliação. Três camadas parecem reunir consenso entre os estudos já apresentados. De reparar que as *neural networks* apresentam maiores dificuldades onde também apresentam dificuldades os modelos físicos. *Genetic algorithm* é um método para a resolução de otimização.

A conjugação do Algoritmo genético com as redes neurais, tem sido aplicados em três áreas [18]:

- pesos estabelecidos para uma estrutura fixa;
- aprender topologias de rede;
- selecionar dados de treinamento e interpretar o comportamento de saída de redes neurais.

Neste estudo, o algoritmo genético foi empregue para determinar as variáveis de entrada e o número de unidades. Cada combinação de variáveis assume uma representação binária enquanto o número de unidades tem base de dez e varia entre um e cinco. É aplicado o algoritmo rápido de Levenberg-Marquardt para treinar as redes neurais.

Temperatura exterior

A previsão da temperatura em ambiente externo (atmosférica) possibilita melhoramentos em sectores chave da economia como agricultura, em que as produções e vegetação pode ser melhorada antecipando acontecimentos de forma a garantir as condições necessárias ao desenvolvimento, monitorização do clima para prever catástrofes naturais, previsão meteorológica, gestão de recursos hídricos, turismo, energia, aviação, telecomunicação, dispersão da poluição, etc. Comparativamente a previsão de temperaturas interiores, a temperatura exterior é menos controlável e as variáveis de *input* podem não ser todas encontradas e ter uma variação não linear. Os modelos estatísticos podem não reproduzir a forte relação não linear entre as variáveis de previsão e por isso em parâmetros atmosféricos (temperatura, velocidade do vento, precipitação, etc,) várias técnicas tem sido aplicadas tais

como a regressão linear, auto regressão. No trabalho [19] é feito um *clustering (k-means)* baseado na lógica *fuzzy* que tem capacidade de acrescentar incerteza na informação, sem entrar em processos físicos e expressões matemáticas complexas.

No estudo foi utilizada a informação dos verões entre 2001-2009 na cidade de Mumbai, India. As variáveis medidas são temperatura máxima, pressão atmosférica, humidade, velocidade do vento e o histórico da temperatura. Estas variáveis pela lógica fuzzy criaram um modelo de previsão da temperatura máxima diária.

O algoritmo *K-means* permite criar grupos, é simples e rápido, basta indicar quantos grupos formar. A informação é agrupada para formar agrupamentos naturais de dados que representam comportamentos comuns.

São agrupados dados de cada variável de entrada com o parâmetro de saída, inicialmente foram obtidos 7 *clusters* mas dada a sua proximidade no final ficaram 3.

Para cada cluster foram feitos diagramas de distribuição de frequências para determinar o alcance e o tipo de pertença que as variáveis vão ter na logica de *fuzzy -Fuzzy Knowledge Base (FKB)* em MATLAB.

FKB é uma extensão do sistema baseado em regras clássica e cria a preciosa incerteza através das regras IF-THEN. É considerada como uma abordagem usada para modelar pela linguagem descritiva com base em lógica fuzzy.

O modelo apresentou uma avaliação de MSE de 3,66 e um erro médio igual a 6,7% comparando com a referência a SVM para previsão de MSE de 7.07 – 7.56 concluindo que a lógica fuzzy pode lidar melhor com as incertezas meteorológicas.

Humidade no solo

Este é o tema que quero desenvolver, dado que o objetivo é criar um modelo de previsão que permite obter qual a humidade no solo nas próximas horas. Como variáveis de entrada terei a temperatura e a humidade relativa do solo e do ambiente da rede de sensores e a juntar a essas variáveis terei a previsão meteorológica da internet

No trabalho [20] é apresentada uma abordagem de *Support Vector Regression (SVR)* que teve origem no *Support Vector Machine (SVM)*, que como é indicado (Vapnik, 1995; e Vapnik Cortes, 1995) revela-se como um poderoso método quando comparado com métodos clássicos e *state-of-the-art*.

Existem várias variáveis necessárias para a previsão de perda de água como a temperatura, velocidade do vento, radiação solar e humidade relativa.

Inicialmente é criado um espaço multidimensional para aplicar o cálculo de uma regressão linear. Neste estudo, é comparado o uso do SVR com as Rede Neural Artificiais (RNA), Autoregressive Movimento Integrado Average (ARIMA), Método Grupo de Manipulação de Dados (GMDH).

Os resultados mostram que as previsões baseadas em SVR têm elevado grau de sucesso, 83% para series univariadas e 97% para multivariadas.

Mas este tipo de abordagem requer entrada especial de formato de dados e de alta velocidade de computadores devido a necessidade de cálculos o que torna uma abordagem pouco aplicada a redes de sensores dado que as decisões têm de ser rápidas e a comunicação de dados é contínua e elevada.

Outra abordagem [21] inicia explicando que existem muitos trabalhos que utilizam apenas índices (índice de precipitação anterior) que refere qual a humidade que o solo “sofreu” e outros híbridos que também utilizam medições no terreno em análise, mas que se torna inviável quando as áreas são grandes e variadas, mas este modelo terá em conta medições, índices de precipitação passada e previstas. O objetivo é classificar entre “seco” e “húmido”, A primeira abordagem é a utilização do k-vizinho mais próximo (KNN) algoritmo, que utiliza todas as observações de treino para efetuar uma classificação, uma análise e previsão. No caso estudado o algoritmo usa a precipitação atual e potencial evapotranspiração e localiza o vizinho mais próximo pela regra de distância de euclidiana. O segundo algoritmo aplicado são as árvores de decisão que dividem os valores em conjuntos por variáveis independentes procurando minimizar a entropia de cada subconjunto maximizando o ganho de informação. As árvores de classificação têm vantagem onde os dados de treino são limitados, (contrariamente ao KNN que necessita de muitos vizinhos para boas classificações) e tem capacidade de ignorar informação que é irrelevante.

O último algoritmo, foi o Neural Network (NN) simples, de camada única, e de forma de *feed-forward*. No processo de reforço das NN é utilizado o algoritmo *AdaBoost* (Freund e Schapire de 1997, descrito por Russell e Norvig, 2010, p. 751), que tem sido utilizado em problemas ambientais, de previsão de pesca, auxiliar na gestão de florestas. É um exemplo de aprendizagem em conjunto, o qual é usado para gerar múltiplas hipóteses (percepções). Permite melhorar o desempenho de qualquer algoritmo de aprendizagem fraca.

O modelo é composto por estes três algoritmos que cada um vota na 'Votação' na predição final ganha a classificação que tiver mais votos.

Valores meteorológicos

O trabalho [22] compara novamente o *Support vector Machine (SVM)*, neste caso o *Least Squares Support vector Machine (LS-SVM)* e as diferentes *Artificial Neural Networks (ANN)*. Mostrando que para previsão de valores o SVM tem mais eficácia, a avaliação foi feita via *cross validation K-fold* e teste de *Kolmogorov-Smirnov*.

ANN vs Modelos Estatísticos

Relativamente as ANN tem dificuldades em lidar com certos sinais não-estacionários (série temporal) e com valores cujo modelo matemático não é linear. As ANN são poderosas quando aplicado a problemas cuja solução requer um conhecimento que é difícil de especificar e que para os quais não existe uma abundância de exemplos pois apenas necessitam de uma quantidade reduzida de informações para a previsão da série temporal. Com base na disposição das séries temporais, os parâmetros de rede interna estão sintonizados usando um algoritmo de ajuste.

A abordagem da rede neural para previsão em series temporais é não paramétrico, logo não precisa saber qualquer informação sobre o processo que gera o sinal, o que contrasta com os processos estatísticos como *auto-regressive (AR)*, *moving-average (MA)*, *AR moving average (ARMA)*, *AR integrated moving-average (ARIMA)* and *Markov chain models nas ANN*, um processo de aprendizagem a partir de *data* treino realiza esta tarefa.

As principais razões para o uso de redes neurais ao invés de análise de séries temporais clássica são:

- Rapidez
- Precisos
- Capazes de lidar com a não-linearidade e dinâmica de entrada de valores
- Previsão paramétricos e não paramétricos.

No entanto as NN são suscetíveis ao comportamento caótico, em formação de mínimos locais e *over-fitting* do problema, pode apresentar dificuldades em conseguir um bom desempenho de previsão.

Um ANN foi utilizado para previsão horária com base em valores médios de temperatura ambiente com 24 h de antecedência, foram utilizados valores horários de um ano completo de temperatura ambiente para treinar um modelo de rede neural.

SVM Vs NN

Como um método alternativo para ANN, o SVM, desenvolvido por Vapnik (1999) pode fornecer um desempenho e previsão melhor no alcance de uma solução de otimização global.

As variáveis não são nem completamente aleatórias nem deterministas e dado que o modelo pode assentar em valores passados, é necessário obter um registo desses dados. Dado que o espaço é limitado o objetivo é ter dados em número adequado e que maximizem o ganho de informação.

O método SVM, que foi sugerido pela primeira vez por Vapnik, recentemente tem sido utilizado numa série de aplicações, tais como classificação, regressão e previsão.

A capacidade de SVM para fazer previsão de regressão não linear faz do modelo bem sucedido na previsão de séries temporais. Para mostrar a eficácia do método LS-SVM, os resultados da previsão foram comparados com quatro redes neurais artificiais perceptron multi-camadas (MLP), função de base radial (RBF), rede neural recorrente (RNN) e probabilística rede neural (PNN).

De acordo com as avaliações o trabalho conclui que os métodos de previsão LS-SVM e PNN fornecem previsões mais precisas do que outras arquiteturas ANN investigados. Obtendo a seguinte classificação, primeiro é o LS-SVM, PNN, PNN, RBF e, finalmente, a MLP.

III. Metodologia de Investigação

O estudo inicia com a implementação, embora simples, de uma rede *wireless* com base em comunicação rádio. A segunda meta foi a recolha de dados e criação automática da base de dados. O terceiro e último objetivo foi estudar a base de dados e aplicar algoritmos de *data mining* para a criação de um modelo de previsão da humidade do solo em estudo para as próximas 4 horas.

Rede de Sensores *Wireless*

Um dos maiores desafios do estudo foi a implementação da rede sem fios em terreno “aberto”. Além da sua implementação foi necessário verificar em condições reais o seu funcionamento (comunicação, resistência à humidade, etc), fiabilidade e consistência dos dados obtidos.

A rede foi implementada sem intervenção de especialista, por essa razão a rede apresenta algumas falhas em termos eletrónicos e de eficiência, mas o principal foco foi montar uma rede o mais simples possível que permitisse o estudo do tema apresentado.

Elementos Hardware e Software

As criações de redes e a utilização de sensores tornaram-se de uso universal, cada pessoa pode ser um criador sem ter grandes conhecimentos de engenharia ou de informática. Nos últimos anos temos observado o aparecimento de *open source hardware* economicamente acessíveis com grande capacidade de armazenamento e processamento como *arduino revolution* para leitores de sensores [1] e todos os seus colones, que permitem a exploração de conceitos como o e-têxtil e o uso de microcontroladores [2] e a fundação *raspberry* que tem lançado vários micro computadores como recentemente lançaram o “*raspberry zero*” [3] ao preço de cinco dólares. Com o lançamento do *hardware*, surge também *open source softwares* que permite criar e manipular as bases de dados como o “R”, *data mining tools*, *packages* para *python*, *Arduino IDE*. Quando juntamos o *open source* com a universalização da internet existe uma partilha de conhecimento através de vários meios web (*youtube*, *blogges*, *Q&A webpages*) o que torna perpétua e acessível a todos os utilizadores. Permite também que quanto mais pessoas exploraram um tema, mais informação existe para ser partilhada. Na nossa rede utilizamos *hardware* e *software* da *arduino.cc* com o auxílio de um PC com sistema operativo *Windows* com *powershell* e *excel*.

Rede

Uma rede é composta por vários nós e ligações entre eles. Existem vários tipos de utilidade a dar aos nós, há nós onde estão instalados os sensores que fazem a leitura e emitem os valores obtidos e outros que agregam a informação e a transmite até atingir o nó central, a estação base. Para simplificar o estudo como demonstra a Figura 3 foi apenas criado um nó que faz leituras de sensores de 30 em 30 minutos e que transmite essa informação para a estação base via comunicação radio. É na estação base (recetor) que a informação é armazenada na base de dados, processada e são aplicadas as técnicas de *data mining*.

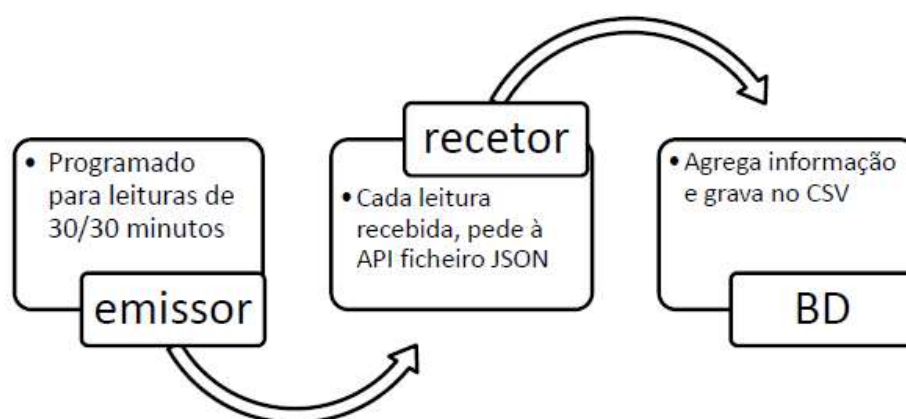


Figura 3 – Workflow da Rede

Composição dos nós da rede

Cada nó é composto por um elemento central, o microprocessador, que tem o processador e a memória onde é alojado o código das ações que queremos executar. É a este microprocessador que são ligados os sensores e o módulo de comunicação. Todo este sistema não funciona autonomamente se não for alimentado, e é aqui que está um dos maiores constrangimentos ao uso de redes de *wireless* [5]. Sem nenhuma ligação física de alimentação e comunicação [11] é necessário calcular quais as necessidades que o nó apresenta e quanto é necessário obter para manter “vivo” durante o tempo que pretendemos. Este é um tema que está em constante investigação, quer da parte de melhores baterias quer na parte de componentes com menor consumo.

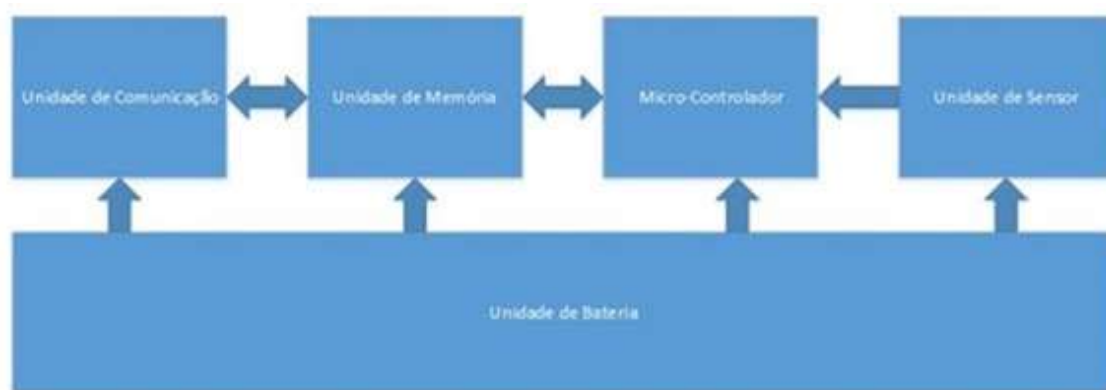


Figura 4 - Composição do Nó

Na rede criada, cada nó utiliza como unidade central um “Arduino uno rev3” que agrega grande parte do circuito, unidades de memória e micro-controlador.

No nó emissor foram acrescentados os sensores, unidade de comunicação e a bateria necessária para ficar autónomo (bateria alimentada por um painel solar). No nó base apenas foi necessário acrescentar a unidade de comunicação e fazer a ligação ao PC com ligação a internet para utilização da API e criar a base de dados.

Sensores

Existem vários tipos de sensores que permitem a baixo custo ter diferentes medições ou reações. Por exemplo, é possível ter uma medição de temperatura, de humidade, de pressão, de proximidade, luminosidade, etc, e também conseguimos ligar um circuito com a presença de alguém.

Os sensores estão em constante evolução com maior precisão e menor consumo, é uma área que está sempre em investigação pois cada vez mais a nossa vida está dependente deles. Monitorização na área da saúde [8], de casas [6], utilização militar, alarmes [7], eficiência energética.

Na rede criada foram utilizados os seguintes tipo sensores (algumas especificações técnicas em anexo):

- Termómetro terra – DS18B20
- Humidade terra – G305
- Temperatura e humidade ar – DHT22 (AM2303)

Ligação e comunicação

Atualmente existem muitas ligações sem fios [4] e cada uma apresenta características que as torna únicas [12], variam muito na conexão, gasto energético, alcance, infraestrutura necessária entre emissor e recetor [13]. Exemplos básicos de comunicação *wireless* é a comunicação por:

- infravermelhos – telecomandos de TV
- *Bluetooth* – nos automóveis para emparelhar equipamentos
- Rádio – rádios que utilizamos nos automóveis
- GSM – comunicação dos nossos telemóveis com as antenas,
- entre outras [14]

mas os principais tipos de *wireless* para uso de automatização na criação de redes são:

- *Radio Frequency*
- *Bluetooth*
- *WI-FI*
- *Wi-Fi HaLow* (03-01-2016) [23]
- *ZigBEE*

No trabalho desenvolvido foi implementada uma rede sem fios com base em radio frequência pois responde às necessidades de distância e fiabilidade. Dado ser alocada a uma área de cultivo de plantas rasteiras e passagem de máquinas, não é viável enterrar nem colocar à superfície os fios para a comunicação e alimentação do nó emissor.

Na Tabela 1 é efetuada a comparação entre redes com fios “*wired network*” e uma rede sem fios “*Wireless network*”. São apresentadas as vantagens e as desvantagens de cada. Na nossa implementação a desvantagem verificada foi a ausência de algumas leituras por erro na comunicação, ou seja, a mensagem que partiu do emissor com os valores dos sensores não chegou ao recetor e por consequência não foi efetuado pedido à API nem feito nenhum registo na base de dados.

Tabela 1 - Diferença entre Rede com fios e Rede sem fios

| Sr. No. | Characteristics | Wired Networks | Wireless Networks |
|---------|---|--|--|
| 1. | Installation | Difficult to moderate (Because More no. of components are used during installation and require cables to be connected to each and every computer in the network) | Easy installation (neat and clean, no untidy cables are used in this) |
| 2. | Visibility Node to Node on same network | All of the nodes on a wired network can hear all other nodes | Many nodes on a wireless network cannot hear all of the other wireless nodes on the same network |
| 3. | Visibility Network to Network | Networks are invisible to other wired networks. The presence of one wired network has no effect on the performance of another wired network | Wireless networks are often visible to other wireless networks. One wireless network can affect the performance of other wireless networks. |
| 4. | Time to installation | More (due to connection of each and every computer in the network) | Less (no untidy cable connections involves in this) |
| 5. | Cost | Less (such Ethernet, cables, switches are not expensive) | More (wireless adapters and access points are quite expensive) |
| 6. | User connectivity | Connectivity is possible only to or from those physical locations where the network cabling extends. | Connectivity is possible beyond the bounds of physical network cabling. |
| 7. | Mobility | Limited (because it operates only on a connected computers linked with the network) | Outstanding (enable wireless user to connect to network and communicate with other users anytime, anywhere) |
| 8. | Reliability | High (Ethernet cables, switches are reliable because manufactures have improving technology over several decades) | Reasonably high(because if the major section like router break down the whole network will be affected) |
| 9. | Speed and Bandwidth | High Up to 100 mbps | Low Up to 54 mbps(depends upon standards 802.11g) |
| 10. | Cables | Ethernet, copper and optical fibers | Works on radio waves and microwaves |
| 11. | Hubs and switches | Need hubs and switches for connections | No need of hubs and switches |
| 12. | Security | Good (by using some software like free wall software etc.) | Weak (because wireless communication signals travel through the air and can easily be intercepted but it can improve by encryption technique) |
| 13. | Types | Local Area Network(LAN) Metropolitan Area network(MAN) Wide Area Network | 1. By Network Formation and Architecture: <ul style="list-style-type: none"> • Infra structure based network • Infra structure less (ad hoc) network |

A nossa rede comunica via radio, com base no módulo NRF24L01 + que é um transmissor de Radio Frequência de 2Mbps de alta potência e ultrabaixo consumo para a banda 2.4GHz ISM (*Industrial, Scientific and Medical* - industrial, científica e médica).

Estação base/ central

É para o nó principal que todos os nós convergem. É neste nó que terá que apresentar maior capacidade de memória e processamento, pois quanto maior a rede, maior o número de dados que chegam. Quanto maior a diversidade dos sensores mais heterogêneos serão os dados. Este nó funciona principalmente como recetor.

Na rede criada o nó central efetivamente é só recetor, e tem o apoio de um computador para ajudar a criar a base de dados e o processamento necessário. Os dados são recebidos e transmitidos via porta serial para o PC.

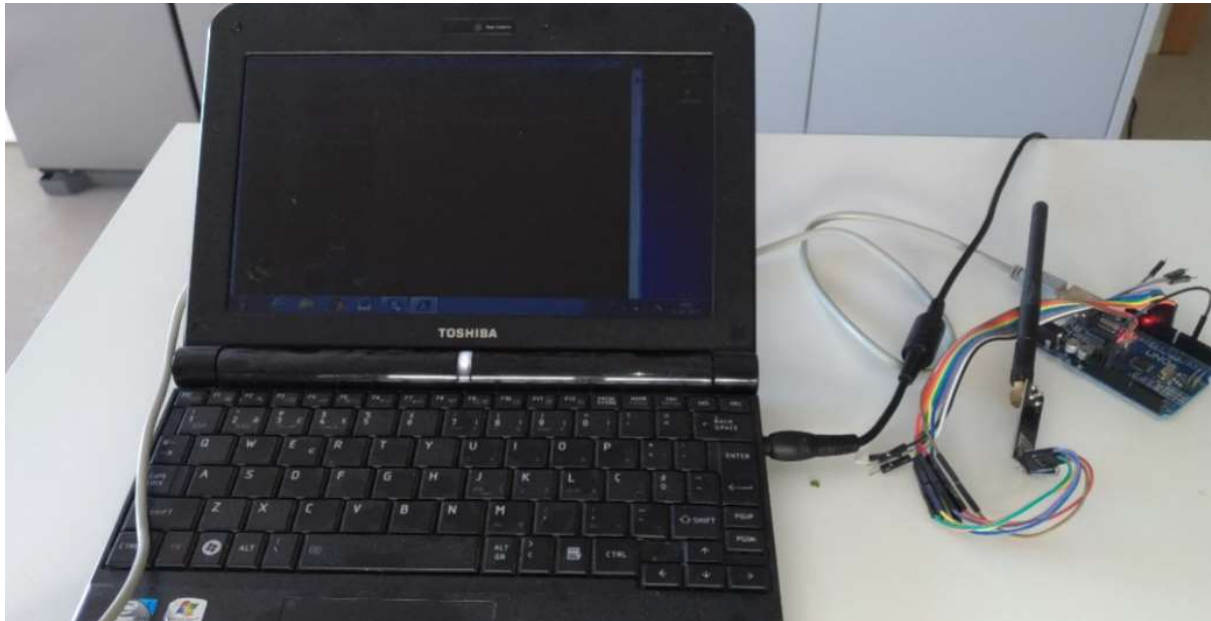


Figura 5 - Nó Base - Recetor e PC

Recolha de dados e criação automática da base de dados

Após escuta da mensagem por parte do *arduino* recetor este comunica por porta serial para o PC. Um exemplo dessa comunicação é visível na figura 6, via IDE de programação do nó foi verificado a informação “escutada” e enviada pela porta COM5, porta que liga o recetor ao PC.

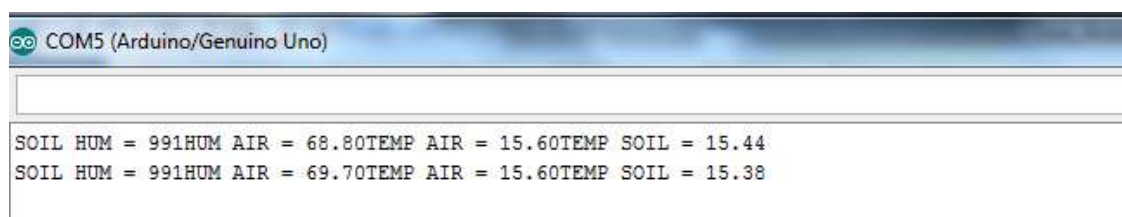


Figura 6 - Dados Recebidos no Recetor e comunicados via serial

Após recebido os dados via porta serial o PC executa a tarefa instruída no *SCRIPT* criado (pode ser consultado nos anexos). É este *SCRIPT* que torna o PC sempre ativo e a verificar se o módulo recetor recebeu alguma informação. Quando recebe uma leitura, o PC executa:

- utiliza o browser para chamar o URL da API (neste URL estão colocadas as coordenadas GPS para receber a informação específica para o local desejado),

- ### Dados da API

No ficheiro recebido, como o da Figura 7, obtemos a informação das previsões para o local escolhido, a informação contém as previsões horárias para as próximas 24 horas e a previsão diária para os próximos 8 dias.

Figura 7 - Dados da API no ficheiro .JSON

As instruções dadas ao PC é para retirar do ficheiro .JSON apenas o necessário para o

estudo para que a base de dados não tenha informação fora do âmbito de estudo. Apenas fica gravado no ficheiro os dados da previsão horária para o momento 0, momento 1, momento 2, momento 3, momento 4. Não sendo possível apresentar todo o ficheiro CSV, é visível na Figura 8 as 4 primeiras variáveis com os dados dos sensores e posteriormente as variáveis fornecidas pela API

| 1 | SoilHum | AirHum | AirTemp | SoilTemp | apparentTemper | cloudCover0 | dewPoint0 | humidity0 | icon0 | ozone0 | p |
|----|---------|--------|---------|-----------|----------------|-------------|-----------|-----------|-------------|--------|---|
| 2 | 609 | 7 | 609 | 72.10 | 20,35 | 0 | 8,43 | 0,46 | clear-night | 318,68 | |
| 3 | 609 | 70.40 | 17.70 | 19.62 | 20,31 | 0 | 8,39 | 0,46 | clear-night | 318,67 | |
| 4 | 639 | 75.60 | 15.60 | 18.06? | 16,22 | 0 | 6,13 | 0,51 | clear-night | 316,37 | |
| 5 | 654 | 8? | 66?? | 671 | 12,87 | 0 | 6,39 | 0,65 | clear-night | 313,95 | |
| 6 | 707 | 92.90 | 9.80 | 10.06 | 9,91 | 0 | 5,3 | 0,73 | clear-night | 313,74 | |
| 7 | 709 | 93.00 | 9.80 | 9.81 | 10,96 | 0 | 6,38 | 0,73 | clear-night | 313,46 | |
| 8 | 712 | 93.10 | 9.70 | 9.56 | 9,28 | 0 | 4,81 | 0,74 | clear-night | 313,17 | |
| 9 | 715 | 92.90 | 9.40 | 9.25 | 11 | 0 | 6,26 | 0,73 | clear-night | 312,79 | |
| 10 | 717 | 92.90 | 9.00 | 8.88 | 8,79 | 0 | 4,18 | 0,73 | clear-night | 312,38 | |
| 11 | 720 | 93.00 | 9.00 | 8.69 | 8,59 | 0 | 4,75 | 0,77 | clear-night | 311,93 | |
| 12 | 722 | 92.90 | 9.10 | ? | 9,26 | 0 | 4,48 | 0,72 | clear-day | 308,33 | |
| 13 | 728 | 93.30 | 8.90 | ng backFa | 11,15 | 0 | 6,26 | 0,72 | clear-day | 308,04 | |
| 14 | 728 | 94.20 | 9.90 | 8.69 | 12,61 | 0 | 7,29 | 0,7 | clear-day | 307,78 | |
| 15 | 727 | 95.80 | 11.70 | 9.63 | 14,84 | 0 | 10,13 | 0,73 | clear-day | 307,54 | |
| 16 | 724 | 96.30 | 13.40 | 10.94 | 16,62 | 0 | 11,09 | 0,7 | clear-day | 307,35 | |
| 17 | 720 | 78.90 | 15.20 | 12.38 | 18,86 | 0 | 12,09 | 0,65 | clear-day | 307,2 | |
| 18 | 718 | 62.20 | 19.60 | 14.06 | 21,37 | 0 | 13,49 | 0,61 | clear-day | 307,12 | |
| 19 | 713 | 53.30 | 24.50 | 19.12 | 21,1 | 0 | 12,72 | 0,59 | clear-day | 307,07 | |
| 20 | 710 | 45.00 | 25.40 | 23.00 | 23,11 | 0 | 15,03 | 0,6 | clear-day | 306,69 | |
| 21 | 710 | 44.70 | 25.50 | 24.19 | 22,38 | 0 | 14,03 | 0,59 | clear-day | 306,64 | |
| 22 | 715 | 38.10 | 25.40 | 23.25 | 22,7 | 0 | 14,21 | 0,59 | clear-day | 306,69 | |
| 23 | 719 | 41.80 | 26.50 | 24.50 | 22,79 | 0 | 12,91 | 0,54 | clear-day | 306,81 | |

Figura 8 - CSV criado com as leituras

Para a aplicação dos modelos foi necessário acrescentar mais umas colunas com valores calculados que serão explicados posteriormente.

Tratamento da base de dados

Neste capítulo abordamos as variáveis, com uma breve descrição de cada uma delas, a sua importância para o modelo e as técnicas aplicadas para as trabalhar. Relativamente aos modelos de previsão serão apenas apresentados no capítulo seguinte.

Dados iniciais

No início os dados retirados da rede e que foram guardados em CSV apresentam um total de 4223 leituras, com a primeira leitura a 11 de abril de 2017 às 20:19 e a última leitura dia 12 de agosto de 2017 às 4:08. Fazendo uma conta rápida de 48 leituras por dia, verificamos que houve bastantes falhas/ausências pois devia ter muitas mais leituras.

Variáveis Recolhidas

Na tabela 2 estão descritas as variáveis do *data set* que estão divididas em três grandes

grupos de origem diferente. O grupo inicial com os dados lidos no terreno no nó emissor que totalizam 4 variáveis. O grupo dos dados da API, recolhidas na estação base da rede, apresentam ainda subgrupos com os períodos de previsão horária (N, N+1, N+2; N+3; N+4), para cada hora foram recolhidas 14 variáveis. E o grupo de variáveis acrescentadas para estudo do modelo. Serão apresentados em detalhe cada grupo de variáveis.

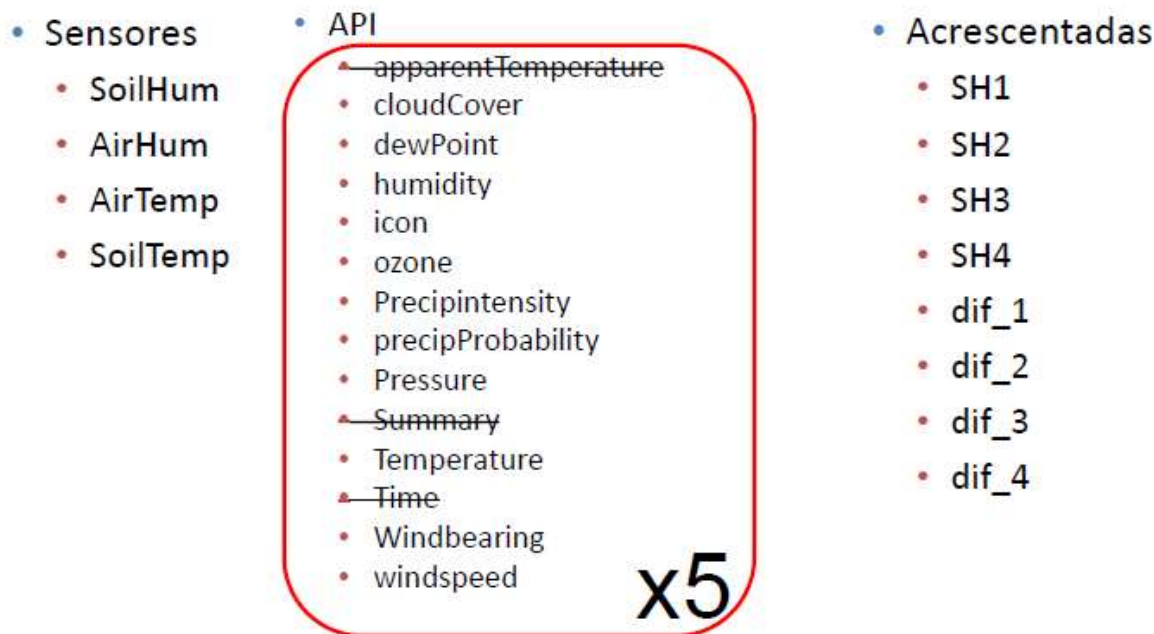


Tabela 2 – Variáveis da base de dados

Dados de sensores no terreno

No estudo, as variáveis recolhidas no nó emissor são as seguintes:

SoilHum – valor dado pelo sensor de humidade do solo – quanto maior o valor menos humidade no solo. No nosso *data set* variou entre 460 e 985.

AirHum – valor dado pelo sensor de humidade do ar a 5cm da superfície da terra do vaso – valor em percentagem.

AirTemp – valor dado pelo sensor de temperatura do ar a 5cm da superfície da terra do vaso – em graus celsius.

SoilTemp - valor dado pelo sensor de temperatura na superfície da terra do vaso – em graus celsius.

Para um estudo inicial às variáveis dos sensores, foram pedidas as estatísticas básicas conforme resumo Tabela 3 e verificamos que existem leituras inconsistentes. A variável da

humidade do ar apresenta um valor máximo de 9420.8% e a temperatura do solo apresenta um valor mínimo de -27.58 graus Celcius temperaturas impossíveis para localização em estudo.

Statistics Table - 2:107 - Statistics

File

| Table "default" - Rows: 67 Spec - Columns: 16 Properties Flow Variables | | | | | | | | | |
|---|----------|--------|---------|---------|--------------|------------|------------|------------|---------------|
| Row ID | S Column | D Min | D Max | D Mean | D Std. de... | D Variance | D Skewness | D Kurtosis | D Overall ... |
| dif_1 | dif_1 | -159 | 655 | 0.26 | 26.253 | 689.234 | 15.787 | 298.678 | 648 |
| SoilHum | SoilHum | 460 | 985 | 783.624 | 116.264 | 13,517.278 | -0.286 | -0.805 | 1,954,359 |
| AirHum | AirHum | 19 | 9,420.8 | 72.98 | 188.753 | 35,627.864 | 48.76 | 2,415.712 | 182,011.63 |
| AirTemp | AirTemp | 8.2 | 43.7 | 21.127 | 6.968 | 48.556 | 0.766 | -0.083 | 52,690.11 |
| SoilTemp | SoilTemp | -27.58 | 48.63 | 21.519 | 7.925 | 62.799 | 0.762 | 0.709 | 53,667.22 |

Tabela 3 - Estatísticas de Variáveis de Sensores

Estas inconsistências foram corrigidas seguindo um critério de média aritmética entre o valor registado na leitura anterior e a seguinte ao erro. Aachamos que é um método que tem pouco impacto de erro na base de dados dado que as variações entre leituras são baixas e as leituras muito próximas uma da outra.

Dados recebidos via API

Outro grupo de variáveis são as “recebidas” via ficheiro da API e agregadas à base de dados via *script*. Este grupo apresenta um conjunto de variáveis multiplicada para cada unidade de tempo (variáveis para o tempo presente, para mais uma hora, para mais duas horas, para mais três horas e para mais quatro horas) foi acrescentado um dígito ao nome da variável para identificar o tempo (0,1,2,3,4). As variáveis para o momento 0 são:

apparentTemperature0 – temperatura aparente em graus celsius

cloudCover0 – valores entre 0 e 1 que representa a percentagem do céu ocupada por nuvens.

dewPoint0 - temperatura atmosférica em graus celsius abaixo da qual as gotículas de água começam a condensar e o orvalho pode se formar.

humidity0 - valores entre 0 e 1 da humidade relativa.

icon0 – imagem que resume a informação recebida, pode ter valores de: *clear-day*, *clear-night*, *rain*, *snow*, *sleet*, *wind*, *fog*, *cloudy*, *partly-cloudy-day*, ou *partly-cloudy-night*.

ozone0 - A densidade atmosférica do ozono em unidades de Dobson.

precipIntensity0 - A intensidade (em polegadas de água líquida por hora) de precipitação que ocorre num determinado momento. Esse valor é condicionado/relacionado à probabilidade (isto é, assumindo que ocorre qualquer precipitação)

precipProbability0 – Probabilidade de ocorrer precipitação.

pressure0 – A pressão do ar do nível do mar em milibares.

summary0 - imagem que sumariza a informação recebia, mais resumido que o “*icon*”

temperature0 – temperatura do ar em graus celcius

time0 – registo da data e hora

windBearing0 - A direção em que o vento sopra em graus, com norte a 0 ° e progredindo no sentido horário

windSpeed0 – velocidade do vento medido em metros por segundo

Posteriormente foram feitas duas análises às variáveis obtidas. A primeira análise apenas por observação, descartamos as seguintes variáveis *apparentTemperature*, *summary* e *time* por apresentar valores idênticos a outras variáveis.

apparentTemperature – verificamos que é idêntica a outra temperatura contida no ficheiro com o nome de variável *temperature*. É uma redundância estar no *data set*.

summary – contem o nome da imagem, das poucas variáveis do tipo *string*, que descreve o tempo meteorológico mas deixamos ficar a variável *icon* (igualmente do tipo *string*) que apresenta um maior detalhe.

time – foi mantido o *time0* que marca o *timestamp* da leitura e do pedido de dados mas as restantes variáveis (*time1*, *time2* *time3* e *time4*) foram retiradas dado que são sempre x horas posteriores a data *time0*, não apresenta mais informação.

Numa segunda análise foi verificado o impacto das variáveis no erro dos modelos e todos melhoram com o retirar das variáveis **cloudCover** e **dewPoint**.

Dados acrescentados

Após analisar os dados, com o objetivo de criar um modelo de previsão, foi necessário criar mais algumas colunas na base de dados. À base de dados foram acrescentadas 8 colunas, 4 com o cálculo de diferença entre as previsões horárias e outras quatro com as leituras que queremos prever.

Variáveis que queremos prever:

SH1 – valor da humidade do solo verificado 2 leituras posteriores ao estado atual

SH2 – valor da humidade do solo verificado 4 leituras posteriores ao estado atual

SH3 – valor da humidade do solo verificado 6 leituras posteriores ao estado atual

SH4 – valor da humidade do solo verificado 8 leituras posteriores ao estado atual

Dado serem leituras de meia em meia hora foi necessário “saltar” duas para completar uma hora. Relembro que relativamente às falhas de leitura não foi efetuado nenhum procedimento de correção pelo que quando saltamos duas leituras pode ser superior a 1 hora mas foi considerada como sendo leitura de uma hora depois.

Variáveis com as diferenças:

dif_1 – diferença entre o valor de humidade do sensor e o valor SH1

dif_2 – diferença entre o valor de SH1 e o valor SH2

dif_3 – diferença entre o valor de SH2 e o valor SH3

dif_4 – diferença entre o valor de SH3 e o valor SH4

O cálculo destas diferenças permite identificar onde surgiram acontecimentos marcantes ao nível da humidade (acontecimentos “aleatórios” como a rega sem nenhuma regra e a chuva) e retirá-los do *data* treino para criar um modelo mais ajustado e focado apenas no comportamento do solo.

De realçar que para um trabalho mais profundo e abrangente a base de dados deveria ter uma coluna com a identificação do acontecimento marcante.

Seleção dos registos da base de dados

Como já foi anteriormente referido, a comunicação rádio necessitou de ser ajustada para diminuir a existência de ausência de leituras e de erros nas leituras recebidas. Na tabela 4 é apresentado o resultado nos erros do ajustamento efetuado no início de maio com a introdução do já mencionado *reset*, os erros de leitura abrandaram após 01-05-2017 às 07:43.

Tabela 4 - Erros encontrados na comunicação

| Rótulos de Linha | Erros de leitura |
|--------------------|------------------|
| Abril | 43 |
| Maio | 15 |
| 01-05-2017 00:11 | 1 |
| 01-05-2017 00:44 | 1 |
| 01-05-2017 01:16 | 1 |
| 01-05-2017 01:48 | 1 |
| 01-05-2017 02:20 | 1 |
| 01-05-2017 02:53 | 1 |
| 01-05-2017 03:25 | 1 |
| 01-05-2017 03:57 | 1 |
| 01-05-2017 04:29 | 1 |
| 01-05-2017 05:02 | 1 |
| 01-05-2017 05:34 | 1 |
| 01-05-2017 06:06 | 1 |
| 01-05-2017 06:38 | 1 |
| 01-05-2017 07:11 | 1 |
| 01-05-2017 07:43 | 1 |
| Total Geral | 58 |

Relativamente à quantidade de ausências de leituras verificamos uma queda de abril até junho conforme o apresentado na Tabela 5.

Tabela 5 - Leituras em falta

| Rótulos de Linha | Ausência de leitura |
|--------------------|---------------------|
| Abril | 356 |
| Maio | 167 |
| Junho | 58 |
| Julho | 419 |
| Agosto | 214 |
| Total Geral | 1214 |

O mês de julho apresenta um elevado número de ausências, o detalhe é espelhado na Tabela 6 onde as ausências estão ordenadas por conjunto de faltas consecutivas. Após análise foi retirada a conclusão que este número anormal de faltas teve origem na falta de bateria no nó emissor por duas vezes, perdendo cerca de 100 leituras em cada.

Face a este constrangimento por opção não foi efetuada nenhum ajuste na base de dados, dado que não havendo receção de leitura via radio o *script* não pede o ficheiro a API e por isso é uma total ausência de dados. De verificar que as maiorias das falhas tiveram início na madrugada falhando até ao início da tarde.

Tabela 6 - Leituras em falta - detalhe julho

| Time Stamp | Ausência de leitura |
|------------------|---------------------|
| Julho | 419 |
| 18-07-2017 03:18 | 103 |
| 24-07-2017 13:33 | 100 |
| 20-07-2017 19:25 | 25 |
| 17-07-2017 03:38 | 24 |
| 16-07-2017 03:35 | 22 |
| 15-07-2017 03:53 | 21 |
| 14-07-2017 04:00 | 20 |
| 13-07-2017 04:17 | 20 |
| 11-07-2017 03:56 | 20 |
| 12-07-2017 04:10 | 20 |
| 02-07-2017 12:18 | 17 |
| 10-07-2017 05:59 | 17 |
| 01-07-2017 13:35 | 4 |
| 07-07-2017 18:33 | 4 |
| 18-07-2017 01:41 | 1 |
| 13-07-2017 18:17 | 1 |

Para o estudo foram consideradas as leituras de maio (depois de 01-05-2017 07:43) até julho (até 24-07-2017 13:33). Foi descartado o mês de abril e de agosto. O de abril por apresentar muitos erros e o fim do mês de julho e o mês de agosto por já não apresentar relevância para o estudo por desleixo na humidade da terra que não teve variações por falta de rega e ausência de chuva. O *data set* apresenta então 3.198 registos.

Após análise dos aumentos positivos da humidade nos 3.198 registos, foi retirada a conclusão que para a criação do modelo preditivo era importante retirar as linhas do *data set* que tinham variações positivas na humidade superior a 50 unidades. O valor (50 unidades) do critério foi determinado pelo estudo do comportamento da humidade face à rega efetuada, foi considerado que é um valor que já significa uma variação significativa com intervenção externa. Este procedimento foi adotado para o modelo de aprendizagem e para o de previsão, pois evitamos que o modelo tenha em consideração comportamentos externos aleatórios e o que tente aceitar no que é aleatório. Este processo foi efetuado para as quatro colunas acrescentadas, dif_1, dif_2, dif_3 e dif_4, conforme demonstra a Figura 9 do KNIME. No gráfico 1 é apresentado o comportamento que queremos que o modelo evite aprender com este processo, as leituras obtiveram a humidade de 788 às 09-05-2017 - 21:06:05 depois verificamos que duas horas depois SH2 o índice caiu para 643, possivelmente devido a uma chuva ou uma rega ligeira. Dado que o objetivo é obter 4 previsões foi necessário neste caso retirar da base de dados de aprendizagem os registos em que esta variação esta presente. Foram por isso eliminadas as duas seguintes e a

leitura anterior (momentos que esta variação ocorreu em dif_1, dif_3 e dif_4). Em toda a base de dados este procedimento retirou 140 registos, passando dos 3198 para 3058.

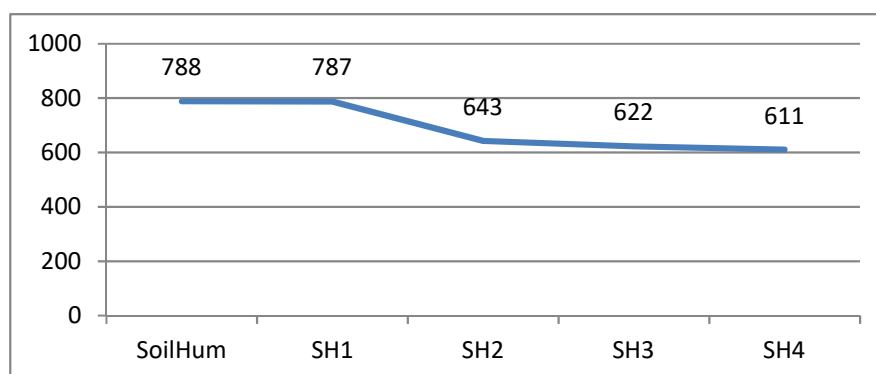


Gráfico 1 – Quebra dos valores observados com acontecimento marcante SH1 para SH2 (dif_2)

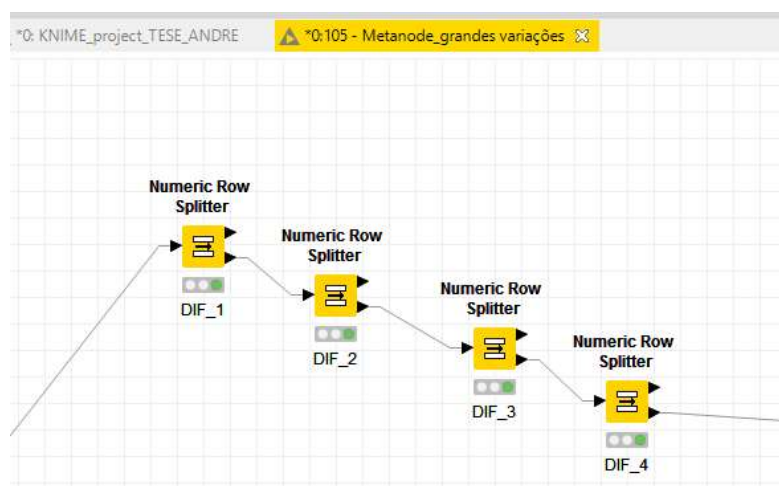


Figura 9 - MetaNode para retirar as grandes variações negativas

Após obter o *data set* para o estudo foi necessário dividir em dados de treino e dados de teste, a base de dados ficou com as características resumidas na Tabela 7.

Tabela 7 - DATA SET

| | DE | A | TOTAL DE REGISTOS | | DIF |
|-------------|------------|------------|----------------------|---|------|
| | | | SEM RETIRAR REGISTOS | SEM GRANDES VARIAÇÕES POSITIVAS (50 UNIDADES) | |
| DATA TREINO | 01-05-2017 | 31-05-2017 | 1195 | 1135 | -60 |
| DATA TESTE | 01-06-2017 | 24-07-2017 | 2003 | 1923 | -80 |
| TOTAL | | | 3198 | 3058 | -140 |

IV. Análise de Dados

Este capítulo está reservado à aplicação dos modelos de previsão ao *data set*. Foi criado um *workflow* no KNIME para criar o modelo de previsão que tem 4 outputs o SH1, SH2, SH3 e SH4. Na Figura 10 está demonstrado o *workflow* mas apenas está visível os nós de previsão para a variável SH4. Serão apresentados também neste capítulo os resultados da avaliação dos modelos para cada período que queremos prever.

O capítulo está dividido em 4 partes, as três primeiras partes é apresentado cada modelo aplicado e uma última parte com os resultados obtidos e alguns estudos específicos com os modelos de melhor desempenho.

Foi efetuado um ensaio preliminar para verificar se o modelo com toda a informação (de todos os tempos em estudo da API) funcionava melhor do que apenas o do tempo que queremos prever e concluí que o erro desce significativamente com o uso de apenas dos dados dos sensores e a previsão da API para esse determinado tempo.

Para ser possível comparar os resultados dos modelos os dados foram normalizados. Em todos os resultados apresentados foi efetuada a comparação com o uso de dados da API e comparado com a utilização apenas de dados lidos pelos sensores.

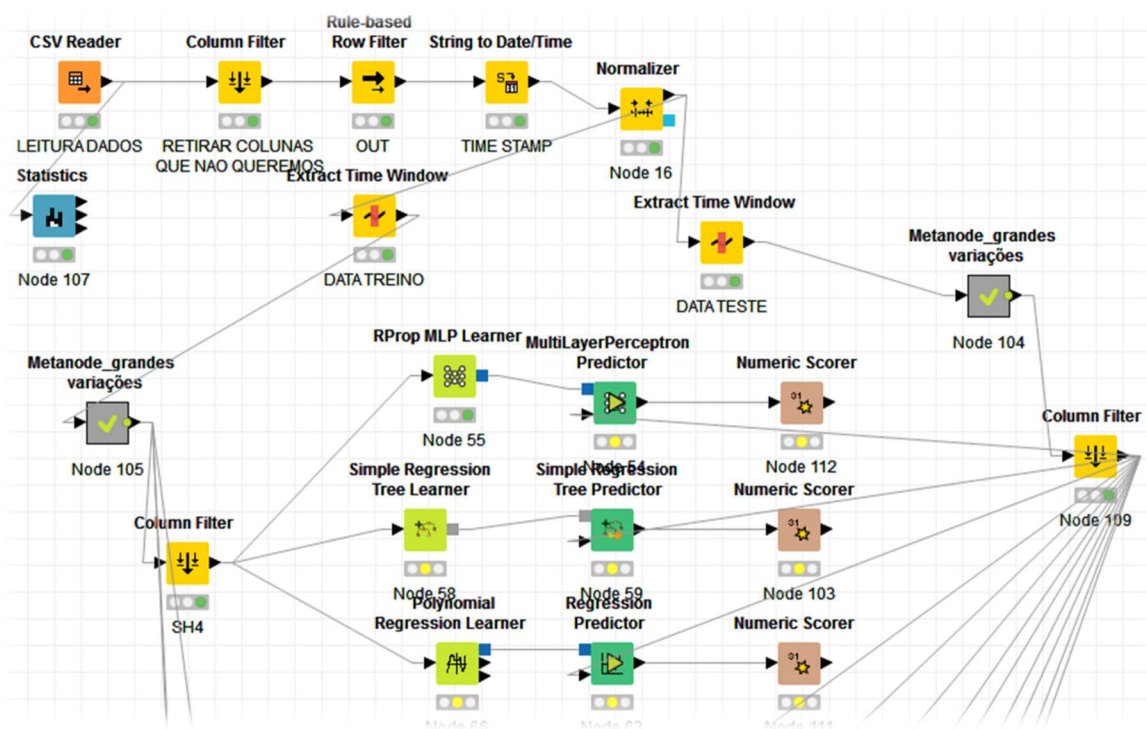


Figura 10 - WORKFLOW do KNIME para aplicação de Modelos para Variável SH4

Multilayer perceptron/ Perceptron Multicamada

Feitos alguns testes e seguindo orientações de alguns dos estudos referenciados anteriormente, foi escolhido, para formar o algoritmo de aprendizagem, apenas uma camada escondida com três neurónios conforme demonstrado na Figura 11.

O resultado obtido no modelo de aprendizagem pode ser estudado pela análise aos coeficientes que compõe o algoritmo, como exemplo na Figura 12 é possível verificar os valores obtidos no modelo para previsão de SH4 com a informação dos sensores e da API para o momento 4.

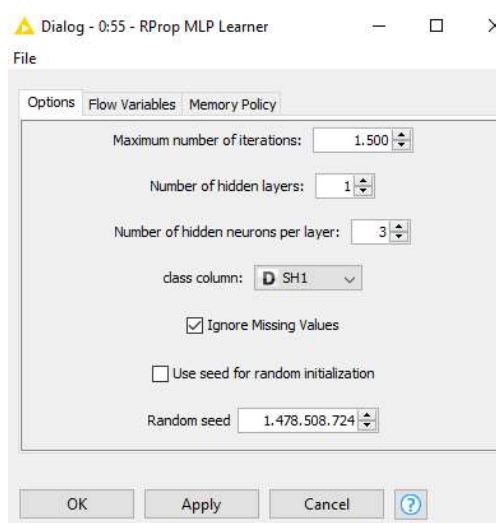


Figura 11 - Configuração do Nó de aprendizagem MLP no KNIME

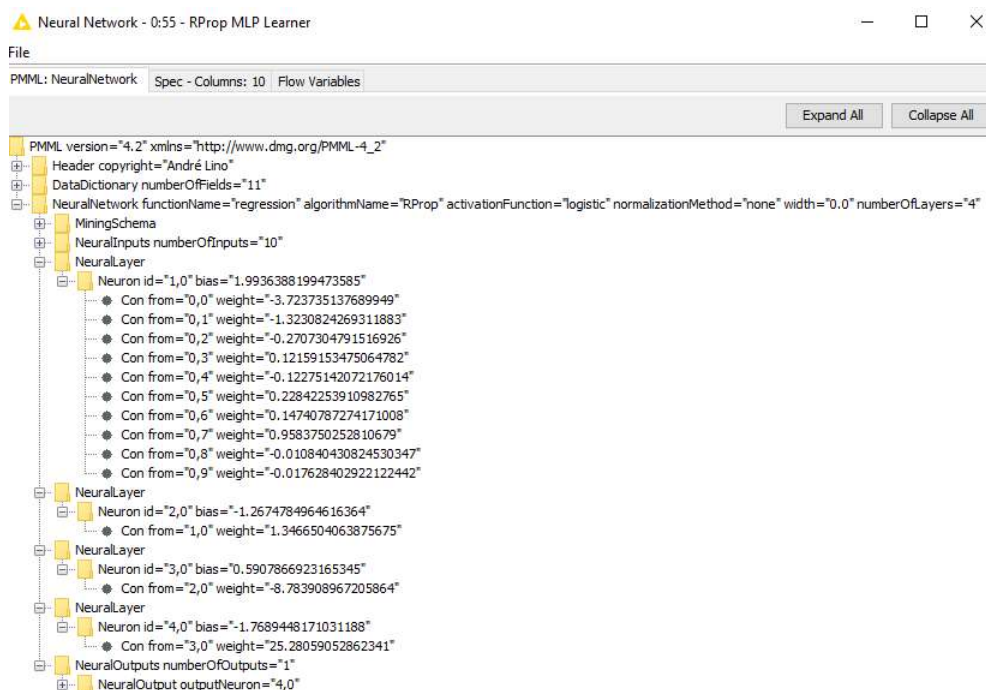


Figura 12 - Valores do algoritmo MLP para a variável SH4

Regression trees / Árvores de Regressão

Este modelo apresenta uma menor capacidade de prever com exatidão, como mostra a Figura 13 os valores variam em degrau.

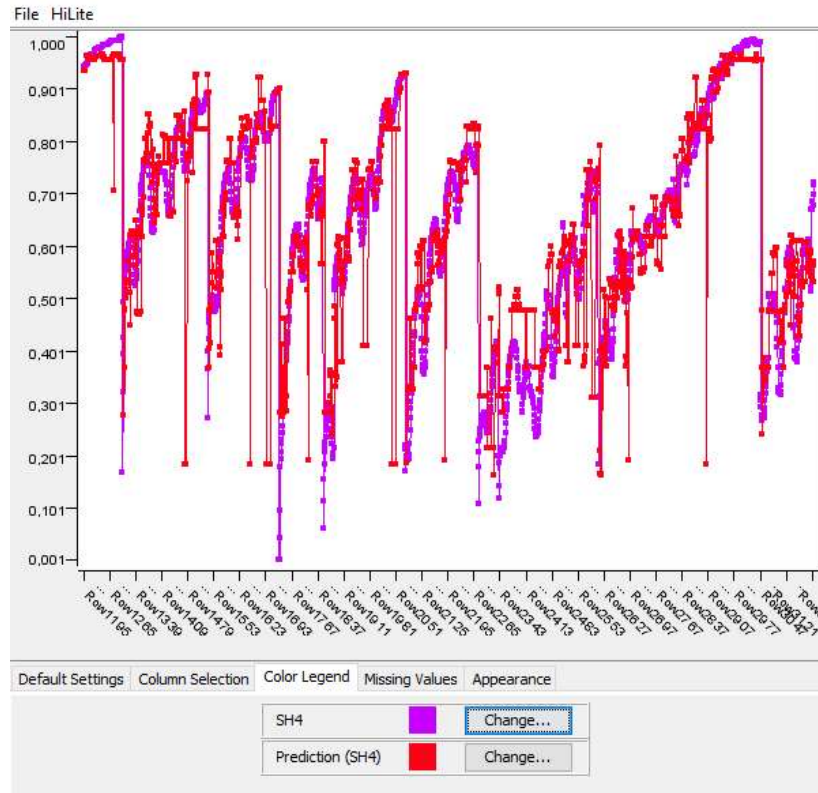


Figura 13 - Gráfico KNIME da variável SH4 e da Previsão SH4 do modelo Regression Trees

Para consultar a árvore de regressão criada, no KNIME é possível verificar que a árvore vai crescendo até enquadrar todos os valores, é apresentado na figura 14 apenas o nó central.

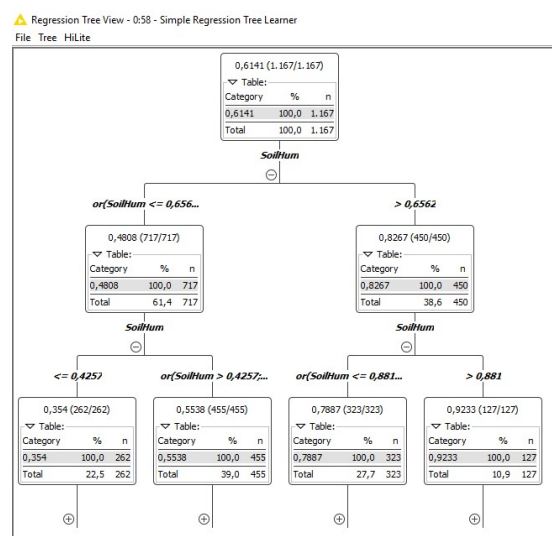


Figura 14 - Desenho do nó da árvore para a variável SH4

Polynomial Regression / Regressão Polinomial

Para determinar que grau de polinômio escolher para aprendizagem do modelo, efetuaram-se testes fazendo variar o grau e analisado o erro pelo MSE. Conclui-se que os resultados melhoraram quando utilizado ao polinômio de segundo grau e o erro aumentava significativamente quando passado do terceiro para o quarto grau.

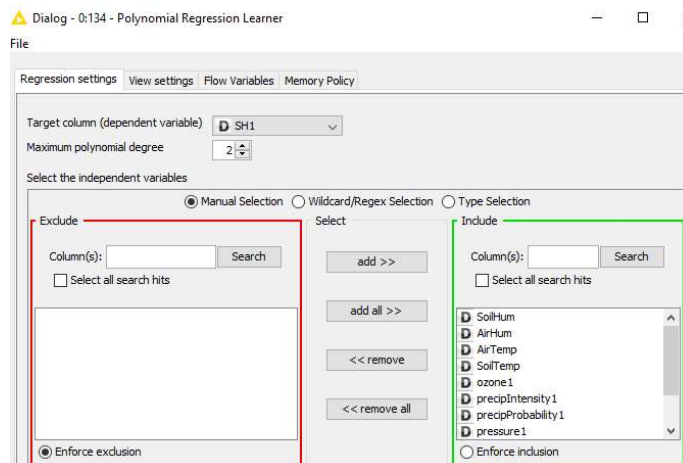


Figura 15 - Configuração do Nó de aprendizagem POLYNOMIAL REGRESSION no KNIME

Como no modelo MLP, na Figura 8 são apresentados os valores que compõe a função polinomial do modelo de previsão para SH4. Modelo aprendido com os dados dos sensores e da API para o momento 4.

Tabela 8 - Valores dos coeficientes do algoritmo POLYNOMIAL REGRESSION para a variável SH4

| Coefficients and Statistics - 0:66 - Polynomial Regression Learner | | | | | | |
|--|-----------------|----------|----------|-------------|-----------|--------|
| File | | | | | | |
| Table "Coefficients and Statistics" - Rows: 21 Spec - Columns: 6 Properties Flow Variables | | | | | | |
| Row ID | S Variable | Exponent | D Coeff. | D Std. Err. | D t-value | D P> t |
| Row1 | SoilHum | 1 | 0.634 | 0.05 | 12.657 | 0 |
| Row2 | SoilTemp | 1 | 0.232 | 0.048 | 4.795 | 0 |
| Row3 | humidity4 | 1 | -0.023 | 0.068 | -0.341 | 0.733 |
| Row4 | ozone4 | 1 | 0.083 | 0.055 | 1.501 | 0.134 |
| Row5 | precipIntens... | 1 | -0.322 | 0.137 | -2.35 | 0.019 |
| Row6 | precipProba... | 1 | 0.163 | 0.048 | 3.377 | 0.001 |
| Row7 | pressure4 | 1 | -0.048 | 0.058 | -0.831 | 0.406 |
| Row8 | temperature4 | 1 | -0.137 | 0.05 | -2.72 | 0.007 |
| Row9 | windBearing4 | 1 | -0.191 | 0.036 | -5.318 | 0 |
| Row10 | windSpeed4 | 1 | 0.056 | 0.039 | 1.44 | 0.15 |
| Row11 | SoilHum | 2 | 0.201 | 0.041 | 4.923 | 0 |
| Row12 | SoilTemp | 2 | 0.028 | 0.061 | 0.455 | 0.649 |
| Row13 | humidity4 | 2 | 0.074 | 0.053 | 1.389 | 0.165 |
| Row14 | ozone4 | 2 | -0.153 | 0.06 | -2.561 | 0.011 |
| Row15 | precipIntens... | 2 | 0.301 | 0.095 | 3.152 | 0.002 |
| Row16 | precipProba... | 2 | -0.119 | 0.08 | -1.491 | 0.136 |
| Row17 | pressure4 | 2 | -0.002 | 0.045 | -0.042 | 0.967 |
| Row18 | temperature4 | 2 | -0.047 | 0.052 | -0.893 | 0.372 |
| Row19 | windBearing4 | 2 | 0.186 | 0.033 | 5.624 | 0 |
| Row20 | windSpeed4 | 2 | -0.023 | 0.042 | -0.546 | 0.585 |
| Row21 | Intercept | 0 | 0.158 | 0.035 | 4.485 | 0 |

Resultados

Neste capítulo são apresentados os resultados dos indicadores de avaliação aos modelos aplicados. Os resultados estão agregados por variável prevista juntando os modelos aplicados.

SH1 - primeira previsão horária da humidade do solo

Foram utilizadas as seguintes variáveis para criar o modelo: SoilHum; AirHum; AirTemp; SoilTemp; humidity1; ozone1; precipIntensity1; precipProbability1; pressure1; temperature1; windBearing1; windSpeed1; SH1

Resultados obtidos nas Tabela 9 , 10 e 11.

Tabela 11 - SH1 no MLP

| | |
|--------------------------|--------|
| File | |
| R ² : | 0,983 |
| Mean absolute error: | 0,02 |
| Mean squared error: | 0,001 |
| Root mean squared error: | 0,028 |
| Mean signed difference: | -0,005 |

Tabela 9 - SH1 no TREES

| | |
|--------------------------|--------|
| File | |
| R ² : | 0,972 |
| Mean absolute error: | 0,023 |
| Mean squared error: | 0,001 |
| Root mean squared error: | 0,036 |
| Mean signed difference: | -0,001 |

Tabela 10 - SH1 no POLYNO.

| | |
|--------------------------|--------|
| File | |
| R ² : | 0,989 |
| Mean absolute error: | 0,015 |
| Mean squared error: | 0 |
| Root mean squared error: | 0,022 |
| Mean signed difference: | -0,002 |

De referir que o valor apresentado do MSE de 0 na Tabela 10 fazendo o calculo o resultado obtido é 0,0004 mas como o KNIME arredonda a 3 casas apresenta o valor de 0.

Posteriormente foi testado o modelo sem dados da API – com as variáveis: SoilHum; AirHum; AirTemp; SoilTemp; SH1

Resultados obtidos nas Tabela 12 , 13 e 14.

Tabela 14 - SH1 sem API no MLP

| | |
|--------------------------|--------|
| File | |
| R ² : | 0,986 |
| Mean absolute error: | 0,018 |
| Mean squared error: | 0,001 |
| Root mean squared error: | 0,025 |
| Mean signed difference: | -0,003 |

Tabela 13 - SH1 sem API no TREES

| | |
|--------------------------|--------|
| File | |
| R ² : | 0,975 |
| Mean absolute error: | 0,023 |
| Mean squared error: | 0,001 |
| Root mean squared error: | 0,034 |
| Mean signed difference: | -0,003 |

Tabela 12 - SH1 sem API no POLYNO.

| | |
|--------------------------|--------|
| File | |
| R ² : | 0,989 |
| Mean absolute error: | 0,014 |
| Mean squared error: | 0,001 |
| Root mean squared error: | 0,022 |
| Mean signed difference: | -0,001 |

Tal como aconteceu com o valor do MSE para o modelo polinomial com dados da API, efetuado o calculo indica 0.0005 que o KNIME arredondou para 0.001 apresentado.

SH2 - segunda previsão horária da humidade do solo

Foram utilizadas as seguintes variáveis para criar o modelo: SoilHum; AirHum; AirTemp; SoilTemp; humidity2; ozone2; precipIntensity2; precipProbability2; pressure2; temperature2; windBearing2; windSpeed2; SH2

Resultados obtidos nas Tabela 15 , 16 e 17.

Tabela 17 - SH2 no MLP

| | |
|--------------------------|--------|
| File | |
| R ² : | 0,964 |
| Mean absolute error: | 0,03 |
| Mean squared error: | 0,002 |
| Root mean squared error: | 0,04 |
| Mean signed difference: | -0,007 |

Tabela 15 - SH2 no TREES

| | |
|--------------------------|--------|
| File | |
| R ² : | 0,938 |
| Mean absolute error: | 0,037 |
| Mean squared error: | 0,003 |
| Root mean squared error: | 0,053 |
| Mean signed difference: | -0,006 |

Tabela 16 - SH2 no POLYN.

| | |
|--------------------------|--------|
| File | |
| R ² : | 0,973 |
| Mean absolute error: | 0,026 |
| Mean squared error: | 0,001 |
| Root mean squared error: | 0,034 |
| Mean signed difference: | -0,008 |

Posteriormente foi testado o modelo sem dados da API – com as variáveis: SoilHum; AirHum; AirTemp; SoilTemp; SH2

Resultados obtidos nas Tabela 18, 19 e 20.

Tabela 18 - SH2 sem API no MLP

| | |
|--------------------------|--------|
| File | |
| R ² : | 0,973 |
| Mean absolute error: | 0,025 |
| Mean squared error: | 0,001 |
| Root mean squared error: | 0,035 |
| Mean signed difference: | -0,002 |

Tabela 20 - SH2 sem API no TREES

| | |
|--------------------------|--------|
| File | |
| R ² : | 0,937 |
| Mean absolute error: | 0,038 |
| Mean squared error: | 0,003 |
| Root mean squared error: | 0,053 |
| Mean signed difference: | -0,003 |

Tabela 19 - SH2 sem API no POLYN.

| | |
|--------------------------|-------|
| File | |
| R ² : | 0,972 |
| Mean absolute error: | 0,025 |
| Mean squared error: | 0,001 |
| Root mean squared error: | 0,035 |
| Mean signed difference: | 0 |

SH3 - terceira previsão horária da humidade do solo

Foram utilizadas as seguintes variáveis para criar o modelo: SoilHum; AirHum; AirTemp; SoilTemp; humidity3; ozone3; precipIntensity3; precipProbability3; pressure3; temperature3; windBearing3; windSpeed3; SH3

Resultados obtidos nas Tabela 21 , 22 e 23.

Tabela 23 - SH3 no MLP




| | | | |
|--|--------|---|---|
|  S. | — |  |  |
| File | | | |
| R ² : | 0,945 | | |
| Mean absolute error: | 0,037 | | |
| Mean squared error: | 0,002 | | |
| Root mean squared error: | 0,049 | | |
| Mean signed difference: | -0,015 | | |

Tabela 21 - SH3 no TREES







| | | | |
|---|-------|---|---|
|  ε | — |  |  |
| File | | | |
| R ² : | 0,895 | | |
| Mean absolute error: | 0,05 | | |
| Mean squared error: | 0,005 | | |
| Root mean squared error: | 0,068 | | |
| Mean signed difference: | 0 | | |

Tabela 22 - SH3 no POLYNO.

| | | | |
|---|-------|---|---|
|  ε | — |  |  |
| File | | | |
| R ² : | 0,961 | | |
| Mean absolute error: | 0,032 | | |
| Mean squared error: | 0,002 | | |
| Root mean squared error: | 0,041 | | |
| Mean signed difference: | -0,01 | | |

Posteriormente foi testado o modelo sem dados da API – com as variáveis: SoilHum; AirHum; AirTemp; SoilTemp; SH3

R Resultados obtidos nas Tabela 24 , 25 e 26.

Tabela 26 - SH3 sem API no MLP




| | | | |
|--|--------|---|---|
|  S. | — |  |  |
| File | | | |
| R ² : | 0,956 | | |
| Mean absolute error: | 0,033 | | |
| Mean squared error: | 0,002 | | |
| Root mean squared error: | 0,044 | | |
| Mean signed difference: | -0,002 | | |

Tabela 25 - SH3 sem API no TREES







| | | | |
|---|-------|---|---|
|  ε | — |  |  |
| File | | | |
| R ² : | 0,897 | | |
| Mean absolute error: | 0,049 | | |
| Mean squared error: | 0,005 | | |
| Root mean squared error: | 0,067 | | |
| Mean signed difference: | -0 | | |

Tabela 24 - SH3 sem API no POLYNO.

| | | | |
|---|-------|---|---|
|  ε | — |  |  |
| File | | | |
| R ² : | 0,954 | | |
| Mean absolute error: | 0,033 | | |
| Mean squared error: | 0,002 | | |
| Root mean squared error: | 0,045 | | |
| Mean signed difference: | 0,001 | | |

SH4 - quarta previsão horária da humidade do solo

Foram utilizadas as seguintes variáveis para criar o modelo: SoilHum; AirHum; AirTemp; SoilTemp; humidity4; ozone4; precipIntensity4; precipProbability4; pressure4; temperature4; windBearing4; windSpeed4; SH4

Resultados obtidos na Tabela 27, 28 e 29:

Tabela 27 - SH4 no MLP

| | |
|--------------------------|--------|
| File | |
| R ² : | 0,928 |
| Mean absolute error: | 0,043 |
| Mean squared error: | 0,003 |
| Root mean squared error: | 0,056 |
| Mean signed difference: | -0,003 |

Tabela 28 - SH4 no TREES

| | |
|--------------------------|-------|
| File | |
| R ² : | 0,858 |
| Mean absolute error: | 0,056 |
| Mean squared error: | 0,006 |
| Root mean squared error: | 0,078 |
| Mean signed difference: | 0,008 |

Tabela 29 - SH4 no POLYNO.

| | |
|--------------------------|--------|
| File | |
| R ² : | 0,949 |
| Mean absolute error: | 0,036 |
| Mean squared error: | 0,002 |
| Root mean squared error: | 0,047 |
| Mean signed difference: | -0,011 |

Posteriormente foi testado o modelo sem dados da API – com as variáveis: SoilHum;

AirHum; AirTemp; SoilTemp; SH4

Resultados obtidos na Tabela 30, 31 e 32:

Tabela 32 - SH4 sem API no MLP

| | |
|--------------------------|-------|
| File | |
| R ² : | 0,912 |
| Mean absolute error: | 0,049 |
| Mean squared error: | 0,004 |
| Root mean squared error: | 0,061 |
| Mean signed difference: | 0,004 |

Tabela 31 - SH4 sem API no TREES

| | |
|--------------------------|--------|
| File | |
| R ² : | 0,867 |
| Mean absolute error: | 0,056 |
| Mean squared error: | 0,006 |
| Root mean squared error: | 0,075 |
| Mean signed difference: | -0,004 |

Tabela 30 - SH4 sem API no POLYNO.

| | |
|--------------------------|-------|
| File | |
| R ² : | 0,936 |
| Mean absolute error: | 0,04 |
| Mean squared error: | 0,003 |
| Root mean squared error: | 0,052 |
| Mean signed difference: | 0,003 |

Resumindo os resultados, e para uma melhor leitura, é apresentado no gráfico 2, 3 e 4 os erros médios quadrados para os diferentes modelos testados.

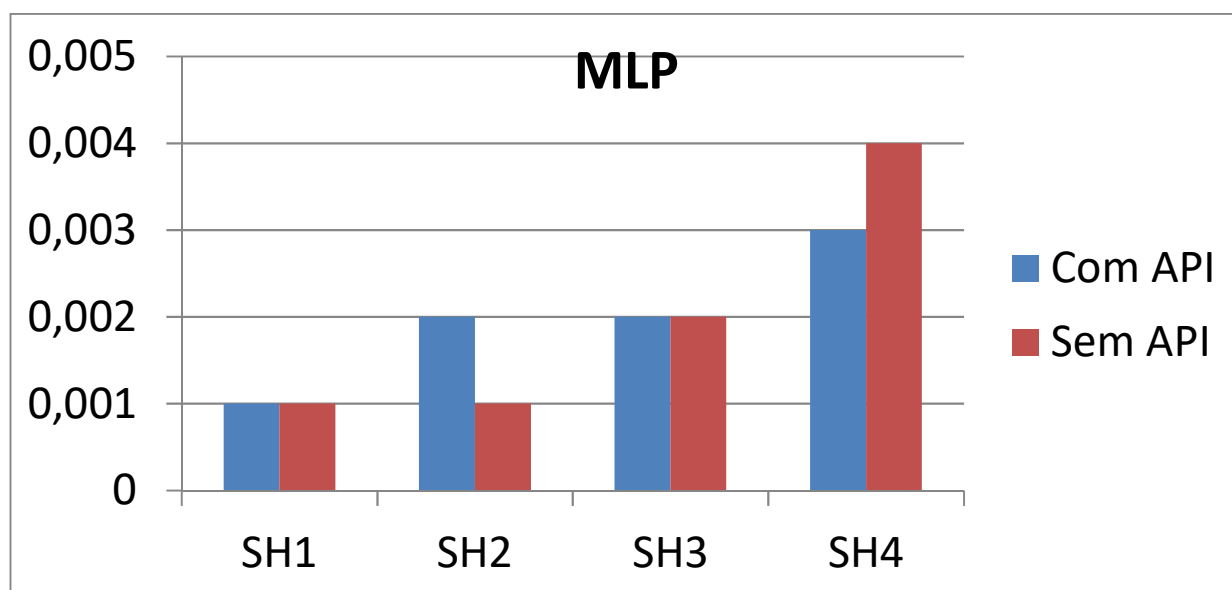


Gráfico 2 - Comparação de resultados do MLP

De verificar que no início o modelo MPL apresenta o mesmo MSE, sendo que nas horas seguintes, sem dados da API, tem um melhor comportamento, mas no momento 4, com uma diferença no MSE de 0,003 para 0,004, apresenta melhor previsão com dados API.



Gráfico 3 - Comparação de resultados do TREES

O modelo Regression Trees apresenta o mesmo resultado de MSE para com e sem dados da API, ou seja, para este modelo os dados não apresentam melhoria na previsão.

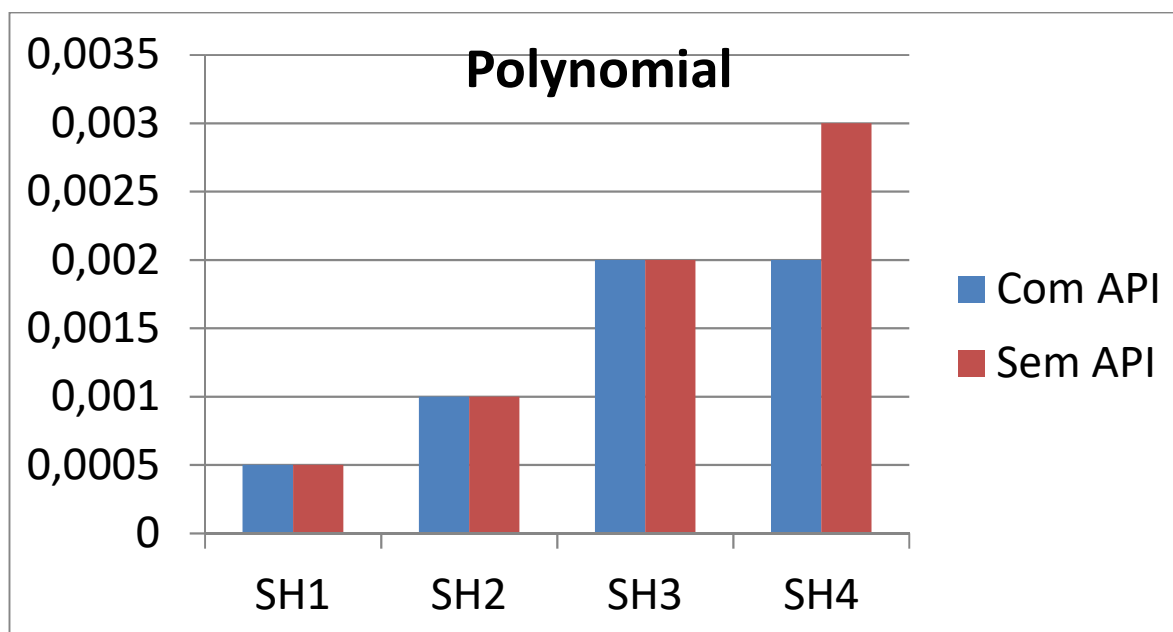


Gráfico 4 - Comparação de resultados do POLYNOMIAL

O modelo Polynomial apresenta ganhos na avaliação com a utilização dos dados da API. Inicialmente iguala o resultado com e sem dados API, depois termina na quarta previsão com uma diferença de MSE de 0.002 para 0.003.

Os melhores modelos de regressão apresentados foram o MLP e o POLYNOMIAL REGRESSION, tendo o último melhor avaliação.

Outros estudos

Após o estudo dos modelos, foram selecionados para mais testes os dois algoritmos com melhor resultado da MSE, deixando o Regression Trees de fazer sentido para o estudo pelo erro apresentado. E dado que o erro vai aumentando com o afastamento do momento da leitura os próximos estudos serão apenas aplicados sobre a previsão de SH4.

O primeiro estudo subsequente efetuado foi para confirmar a importância para a precisão do modelo do processo de retirar as grandes variações. O modelo foi novamente corrido sem este *step* e os resultados obtidos estão nas tabelas 33 e 34. Foi possível assim determinar que estas ações aleatórias de aumento da humidade criam muitos erros na aprendizagem e previsão do modelo. O erro medio quadrado quadruplica.

Tabela 34 - SH4 no MLP sem grandes variações

| | |
|--------------------------|-------|
| File | |
| R ² : | 0,672 |
| Mean absolute error: | 0,062 |
| Mean squared error: | 0,016 |
| Root mean squared error: | 0,125 |
| Mean signed difference: | 0,01 |

Tabela 33 - SH4 no POLYNOMIAL sem grandes variações

| | |
|--------------------------|-------|
| File | |
| R ² : | 0,653 |
| Mean absolute error: | 0,061 |
| Mean squared error: | 0,017 |
| Root mean squared error: | 0,129 |
| Mean signed difference: | 0,013 |

Conclui-se que é melhor retirar as grandes variações, e também que o modelo MPL consegue melhor desempenho com a aleatoriedade.

O segundo estudo subsequente ao estudo inicial foi analisar a alteração dos dados de treino para os meses de maio, junho e os dados de teste para julho. Com esta alteração queremos estudar a importância de aprender com uma base de treino maior. Foram mantidas as condições apresentadas anteriormente (sem retirar as grandes variações). Os resultados são apresentados nas Tabelas 35 e 36.

Tabela 35 - SH4 no MLP aumento data treino

| | |
|--------------------------|--------|
| File | |
| R ² : | 0,822 |
| Mean absolute error: | 0,051 |
| Mean squared error: | 0,007 |
| Root mean squared error: | 0,081 |
| Mean signed difference: | -0,023 |

Tabela 36 - SH4 no POLYNOMIAL aumento data treino

| | |
|--------------------------|--------|
| File | |
| R ² : | 0,838 |
| Mean absolute error: | 0,044 |
| Mean squared error: | 0,006 |
| Root mean squared error: | 0,078 |
| Mean signed difference: | -0,014 |

Com este resultado concluímos que ambos modelos melhoram o seu desempenho na previsão do SH4, por isso é sempre melhor aumentar os dados de treino com as informações que vão sendo testadas.

De seguida foi efetuada a alteração dos dados de treino para o mês de junho e os dados de teste para julho até dia 24, o objetivo foi verificar se apenas com dados mais próximos da nossa previsão conseguimos ou não melhores previsões. Os resultados obtidos são apresentados na tabela 37 e 38, a avaliação dos dois modelos pioram o desempenho principalmente o *polynomial regression*, que tinha apresentado uma melhoria com o aumento da base de treino. Os dois modelos e ficam iguais em MSE.

Tabela 37 - SH4 no MLP

| | |
|--------------------------|--------|
| File | |
| R ² : | 0,8 |
| Mean absolute error: | 0,057 |
| Mean squared error: | 0,007 |
| Root mean squared error: | 0,086 |
| Mean signed difference: | -0,028 |

Tabela 38 - SH4 no POLYNOMIAL
aproximação data treino

| | |
|--------------------------|--------|
| File | |
| R ² : | 0,8 |
| Mean absolute error: | 0,057 |
| Mean squared error: | 0,007 |
| Root mean squared error: | 0,086 |
| Mean signed difference: | -0,028 |

Após estes testes com a alteração dos dados de treino, retiramos a conclusão que o melhor para o modelo é ter os dados mais recentes e trabalhar com dois ou três meses. O modelo deve trabalhar com “*sliding window*”, sendo que cada leitura nova deixa cair a mais antiga.

Posteriormente foi testada a relevância das variáveis dos sensores e qual a sua importância no modelo. Tal como nos últimos testes efetuados foram utilizados os modelos MLP e *POLYNOMIAL REGRESSION* para a previsão da variável SH4, foram retiradas as grandes variações e utilizadas como dados de treino o mês de maio e os meses de junho e julho para dados de teste. Para ambos os modelos foram retiradas as variáveis “Airhum” e “Airtemp” e foram obtidos os resultados da Tabela 39 e 40.

Tabela 40 - SH4 no MLP sem
“Airhum” e “Airtemp”

| | |
|--------------------------|--------|
| File | |
| R ² : | 0,927 |
| Mean absolute error: | 0,043 |
| Mean squared error: | 0,003 |
| Root mean squared error: | 0,056 |
| Mean signed difference: | -0,007 |

Tabela 39 - Tabela 39 - SH4
no POLYNOMIAL sem
“Airhum” e “Airtemp”

| | |
|--------------------------|-------|
| File | |
| R ² : | 0,929 |
| Mean absolute error: | 0,044 |
| Mean squared error: | 0,003 |
| Root mean squared error: | 0,055 |
| Mean signed difference: | -0,02 |

Foi possível concluir que a retirada destas duas variáveis o resultado melhorou para o modelo MPL e teve um pequeno aumento do MSE para a regressão polinomial. Concluimos que é possível simplificar o nó emissor retirando o sensor DHT22, minimizando os gastos e o tempo de leitura.

De seguida foi analisado a importância da outra variável fornecida pelo sensor de temperatura do solo “SoilTemp”. O modelo foi corrido apenas com a variável de sensores “SoilHum” e os dados da API. Foram retiradas as grandes variações e utilizadas como dados de

treino o mês de maio e os meses de junho e julho para dados de teste. A avaliação visível nas Tabelas 41 e 42 permitiu concluir que os resultados obtidos pioram nos dois modelos, mas para um a perda de ganho é insignificativa e no outro modelo a perda é grande.

Tabela 42 - SH4 no MLP sem "Airhum", "Airtemp" e "Soiltemp"

| | |
|--------------------------|-------|
| File | |
| R ² : | 0,922 |
| Mean absolute error: | 0,046 |
| Mean squared error: | 0,003 |
| Root mean squared error: | 0,058 |
| Mean signed difference: | 0 |

Tabela 41 - SH4 no POLYNOMIAL sem "Airhum", "Airtemp" e "Soiltemp"

| | |
|--------------------------|--------|
| File | |
| R ² : | 0,889 |
| Mean absolute error: | 0,057 |
| Mean squared error: | 0,005 |
| Root mean squared error: | 0,069 |
| Mean signed difference: | -0,023 |

Com a retirada desta variável do modelo também pode ser descartado o sensor de temperatura do solo, deixando o nó emissor da rede apenas com o sensor de humidade de solo que indica o valor da variável a prever. Com esta simplificação do nó o modelo que melhor consegue prever a humidade do solo é o MPL.

Foram tentadas outras abordagens que foram descritas na revisão bibliográfica mas não foi obtido nenhum resultado comparável ao que foi obtido. Foi abordado o modelo SVM mas não foi obtido nenhum resultado. Outra abordagem foi a criação de *clusters* pelo método de *k-means*, agregando os valores de humidade em “patamares” e utilizar outros modelos de análise de dados para classificar, mas os resultados foram considerados piores que os obtidos pelo modelo com melhor avaliação (com a vantagem de o modelo prever um número e não uma categoria).

Outra tentativa de melhorar o resultado obtido foi fazer uma “construção temporal”, ou seja, utilizar no modelo de regressão as variáveis das previsões anteriores conforme o esquema apresentado na Figura 16. A ideia deste tipo de construção teve como objetivo inicial tirar partido da boa previsão dos valores para a hora seguinte. Em todos os modelos estudados anteriormente, o MSE para a primeira hora é muito baixo quando comparado com as outras horas. A introdução desta medida foi para dar ao modelo mais variáveis com os valores anteriormente.

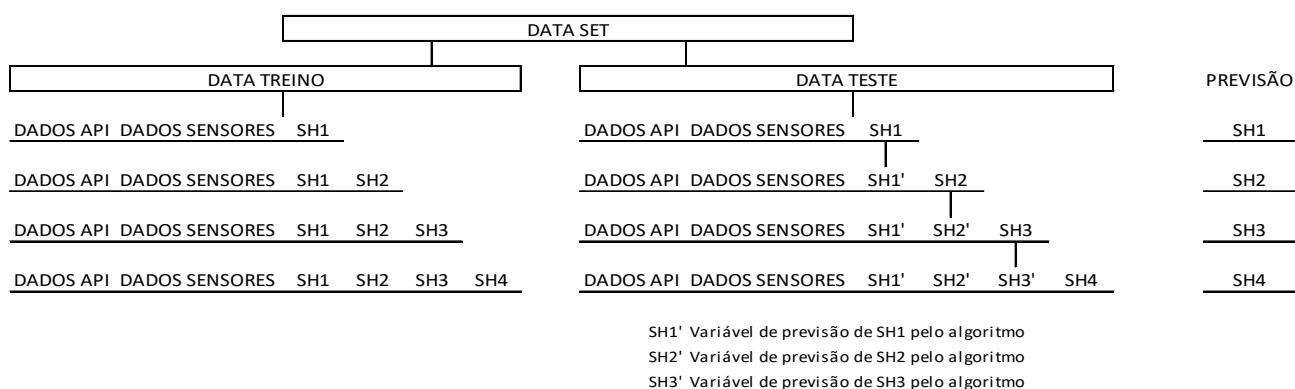


Figura 16 - Variáveis da nova abordagem

O resultado da MSE desta abordagem por modelo está resumido nos gráficos 5, 6 e 7 para as variáveis com dados da API. e nos gráficos 8, 9 e 10 para as variáveis só dos sensores.

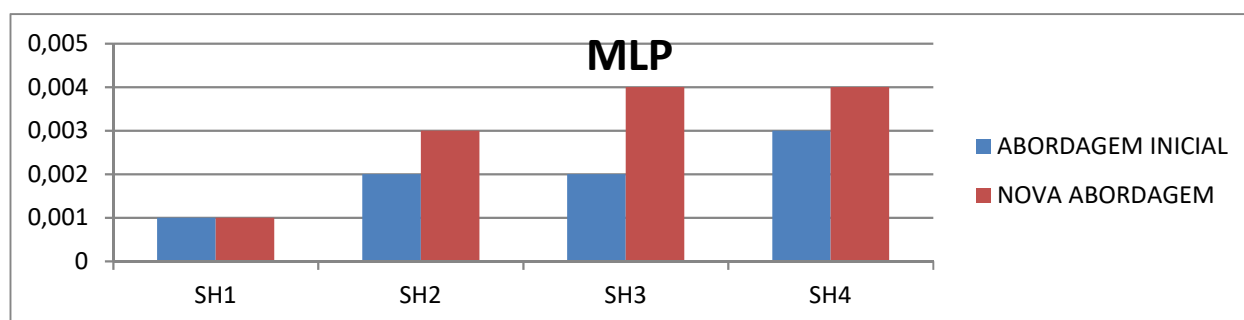


Gráfico 5 - Resumo nova abordagem MLP com dados API

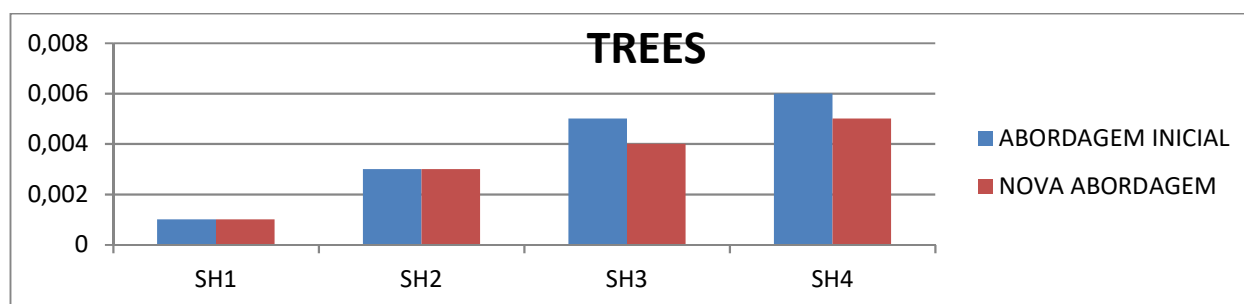


Gráfico 6 - Resumo nova abordagem TREES com dados API

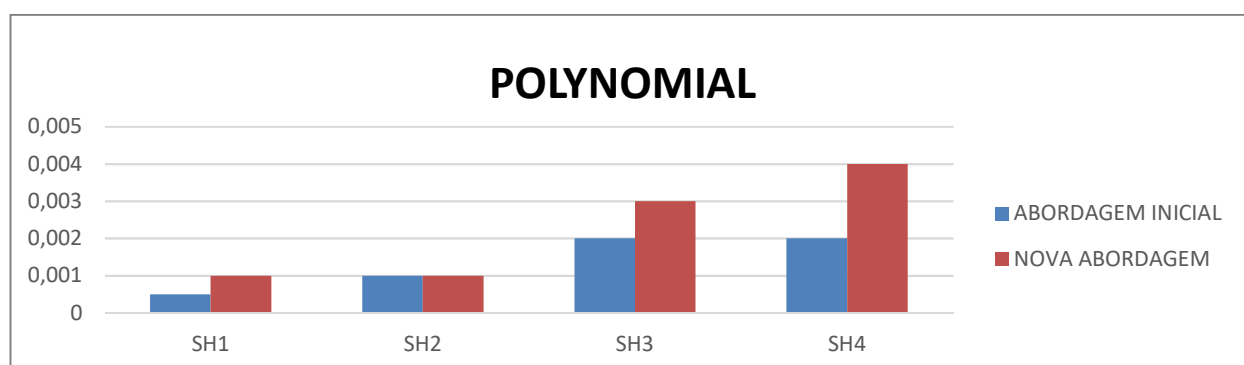


Gráfico 7 - Resumo nova abordagem POLYNOMIAL com dados API

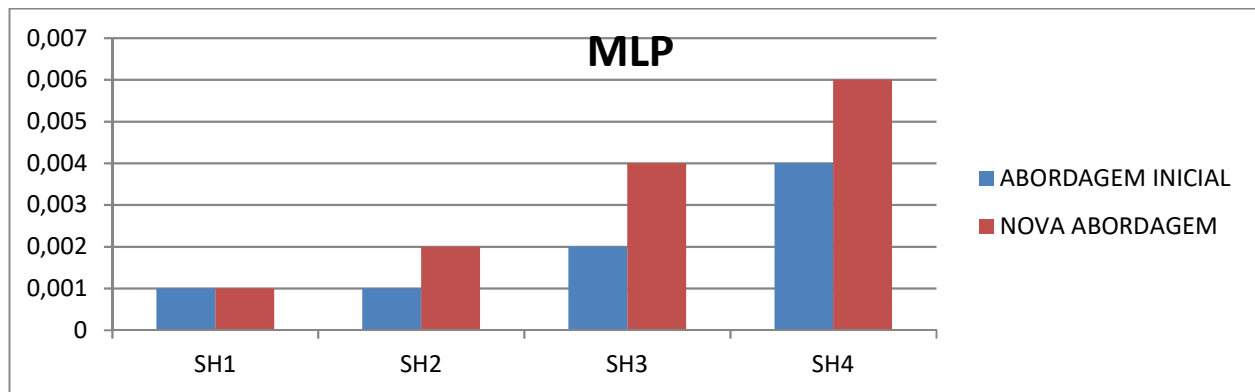


Gráfico 8 - Resumo nova abordagem MLP sem dados API

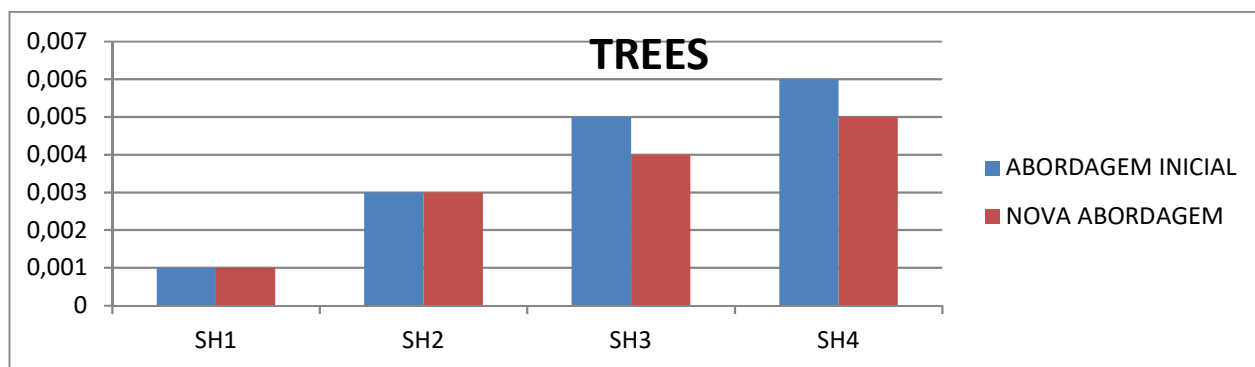


Gráfico 9 - Resumo nova abordagem TREES sem dados API

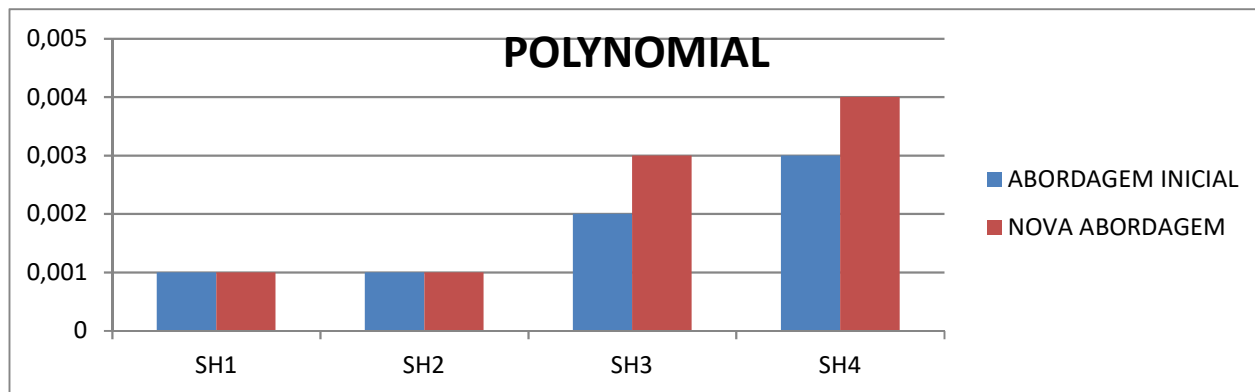


Gráfico 10 - Resumo nova abordagem POLYNOMIAL sem dados API

Nesta abordagem o melhor resultado foi o modelo polinomial sem dados da API mas com um comportamento pior do que o mesmo modelo sem dados da API da abordagem inicial. Os modelos tiveram um desempenho pior exceto as *regression trees*, o que faz sentido dado que este modelo vai categorizando e dividindo tendo como base os valores das variáveis.

V. Conclusões e Pesquisas Futuras

Com este estudo concluímos que com a conjugação de dados reais com informação de *forecast* disponível via API é possível fazer previsões satisfatórias. Foi estudada a capacidade de criar um método de previsão, sem recorrer a dispendiosas estações meteorológicas para obter dados e que a previsão melhora utilizando estes dados “extra” de origem *on-line* quanto mais longínqua for essa previsão. Tendo como base todas as abordagens feitas, modelos aplicados e avaliações MSE e MAE, concluo que nas duas primeiras previsões deve ser utilizado o modelo sem dados da API mas que daí para a frente o modelo deve incluir os dados da API, pois são estes que indicam as condições meteorológicas futuras.

Relativamente ao modelo a aplicar, com as variáveis estudadas, concluímos que o modelo com melhor avaliação foi o *POLYNOMIAL REGRESSION* por ser simples e se adaptar bem as previsões lineares quando foram retiradas as grandes variações. No entanto após simplificação do uso de sensores, ficando apenas o da humidade do solo, o melhor modelo é significativamente o *MLP*. O modelo *REGRESSIONS TREES* não deverá ser descartado para futuros estudos se utilizadas variáveis com categorias (por exemplo como a variável que foi retirada “*ICON*”/“*SUMMARY*” ou introduzir via criação de *clusters* categorias como dia/noite ou quente / frio ou seco/húmido.

Relativamente a pesquisas futuras existem várias abordagens, umas delas com vertentes diferentes de estudo e outras para aplicação do modelo, não devendo nenhum modelo de regressão ser eliminado.

Uma vez que os dados da API têm a informação horária das próximas 24 horas e a informação diária dos próximos 8 dias, uma evolução natural deste estudo será alargar para 8 e 12 horas as previsões e avaliar se as previsões são satisfatórias.

O exemplo de outras vertentes diferentes de estudo, será utilizar o mesmo conjunto de variáveis, mas acrescentar outras com a variação entre as variáveis como taxa de crescimento da temperatura entre o momento 0 e o momento N e verificar se essas variações são mais úteis que os dados e a sua multiplicidade de momentos.

Possibilidade de utilização da sazonalidade do dia, estudar se a hora do dia influencia as próximas humidades, utilizando o sen e cos da hora.

Faltou a este estudo verificar outra API e se traz maior fiabilidade aos resultados.

Exemplo de utilização deste modelo deve ser se este pode servir de base para

aplicação como base inicial de estudo de outro terreno (data treino inicial). Outra aplicação é a criação de um programa de *machine learning* para regadio agrícola. Para isso ser possível é necessário conhecer a cultura que queremos fazer crescer sabendo a humidade que a planta necessita, e dado ser em terreno aberto, o comportamento da terra e a previsão meteorológica em termos de temperatura (geadas, calor extremo, precipitação, etc). Este estudo deve responder ao comportamento da terra e como esta se comporta quando adicionando humidade. Agora deve ser efetuado um estudo semelhante de *machine learning* para prever chuva para ser capaz de saber quando regar e evitar excesso de humidade pois as plantas são mais sensíveis ao excesso de água do que a falta desta.

Se a base de dados tivesse como variável o acontecimento de “rega”, seria interessante juntar à gravação dos dados a decisão de um especialista que indicaria quando regar e a quantidade necessária. Achamos que após uma base de dados alargada, um modelo de classificação conseguiria tomar a decisão de rega e de quantidade de água.

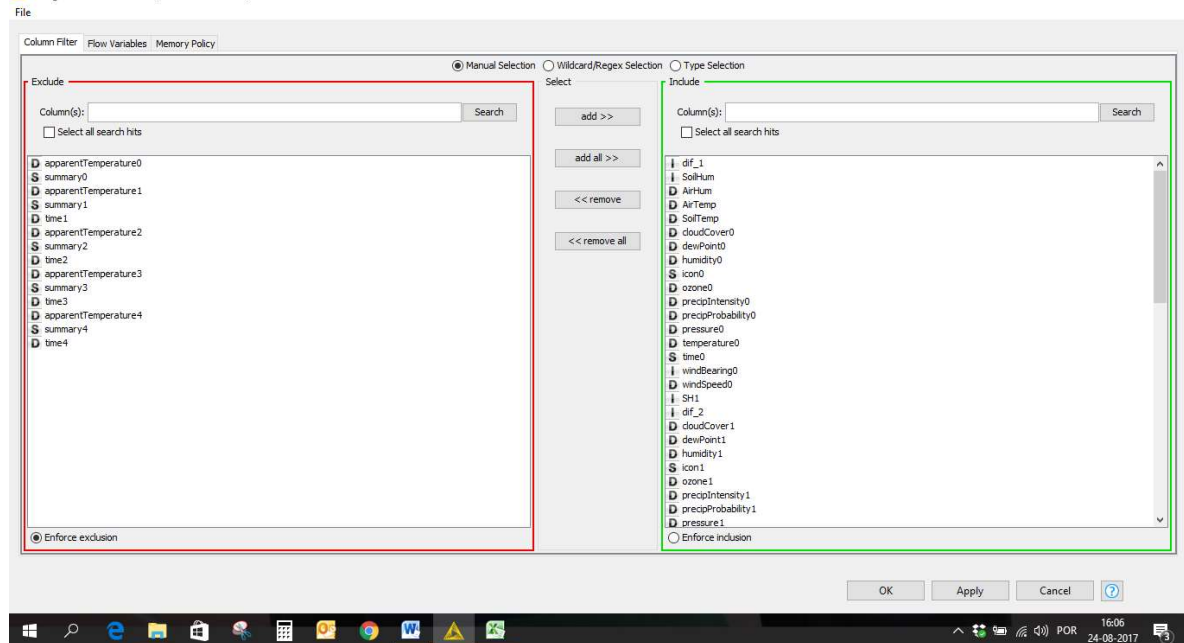
Referências Bibliográfica

1. Bakar Sidik, MA, Arifin Rusli, MQ, Adzis, Z, Buntat, Z, Arief, YZ, Shahroom, H, Nawawi, Z, & Jambak, MI 2015, 'Arduino-Uno Based Mobile Data Logger with GPS Feature', *Telkonnika*, vol. 13, no. 1, pp. 250-259.
2. Buechley, L, & Eisenberg, M 2008, 'The LilyPad Arduino: Toward wearable engineering for everyone', *IEEE Pervasive Computing*, vol. 7, no. 2, p. 12-15. Available from: 10.1109/MPRV.2008.38.
3. 'Raspberry Pi shrinks down to Pi Zero', 2015, *Electronics Weekly*, no. 2650, p. 8.
4. KANTOR, E, LOTKER, Z, PARTER, M, & PELEG, D 2015, 'The Topology of Wireless Communication', *Journal of the ACM*, vol. 62, no. 5, pp. 37-37:32. Available from: 10.1145/2807693.
5. HANGA, C, & FILIP, I 2015, 'ENERGY EFFICIENCY IN WIRELESS SENSOR NETWORKS', *Annals of the Faculty of Engineering Hunedoara - International Journal of Engineering*, vol. 13, no. 4, pp. 267-270.
6. Li, M, & Lin, H 2015, 'Design and implementation of smart home control systems based on wireless sensor networks and power line communications', *IEEE Transactions on Industrial Electronics*, no. 7, p. 4430.
7. Sun, X, Yan, B, Zhang, X, & Rong, C 2015, 'An Integrated Intrusion Detection Model of Cluster-Based Wireless Sensor Network', *PLoS ONE*, vol. 10, no. 10, pp. 1-16. Available from: 10.1371/journal.pone.0139513.
8. Fezari, M, Rasras, R, & Emary, IE 2015, 'Ambulatory Health Monitoring System Using Wireless Sensors Node', *Procedia Computer Science*, vol. 65, no. International Conference on Communications, management, and Information technology (ICCMIT'2015), pp. 86-94. Available from: 10.1016/j.procs.2015.09.082.
9. Fouad, MM, Oweis, NE, Gaber, T, Ahmed, M, & Snasel, V 2015, 'Data Mining and Fusion Techniques for WSNs as a Source of the Big Data', *Procedia Computer Science*, vol. 65, no. International Conference on Communications, management, and Information technology (ICCMIT'2015), pp. 778-786. Available from: 10.1016/j.procs.2015.09.023.
10. Azhar Mahmood, Ke Shi, Shaheen Khatoon, and Mi Xiao, "Data Mining Techniques for Wireless Sensor Networks: A Survey," *International Journal of Distributed Sensor Networks*, vol. 2013, Article ID 406316, 24 pages, 2013. doi:10.1155/2013/406316
11. Navpreet Kaur and Sangeeta Monga "COMPARISONS OF WIRED AND WIRELESS NETWORKS: A REVIEW" *International Journal of Advanced Engineering Technology*, E-ISSN 0976-3945
12. PAL, S 2015, 'EVALUATING ROUTING PERFORMANCES IN MOBILE OPPORTUNISTIC NETWORKS USING DIFFERENT WIRELESS COMMUNICATION TECHNIQUES', *Annals of the Faculty of Engineering Hunedoara - International Journal of Engineering*, vol. 13, no. 4, pp. 57-62.
13. Ramiro Jordan and Chaouki T. Abdallah, "Wireless Communications and Networking: An overview"
14. V.Abinayaa*, Anagha Jayan, "Case Study on Comparison of Wireless Technologies in

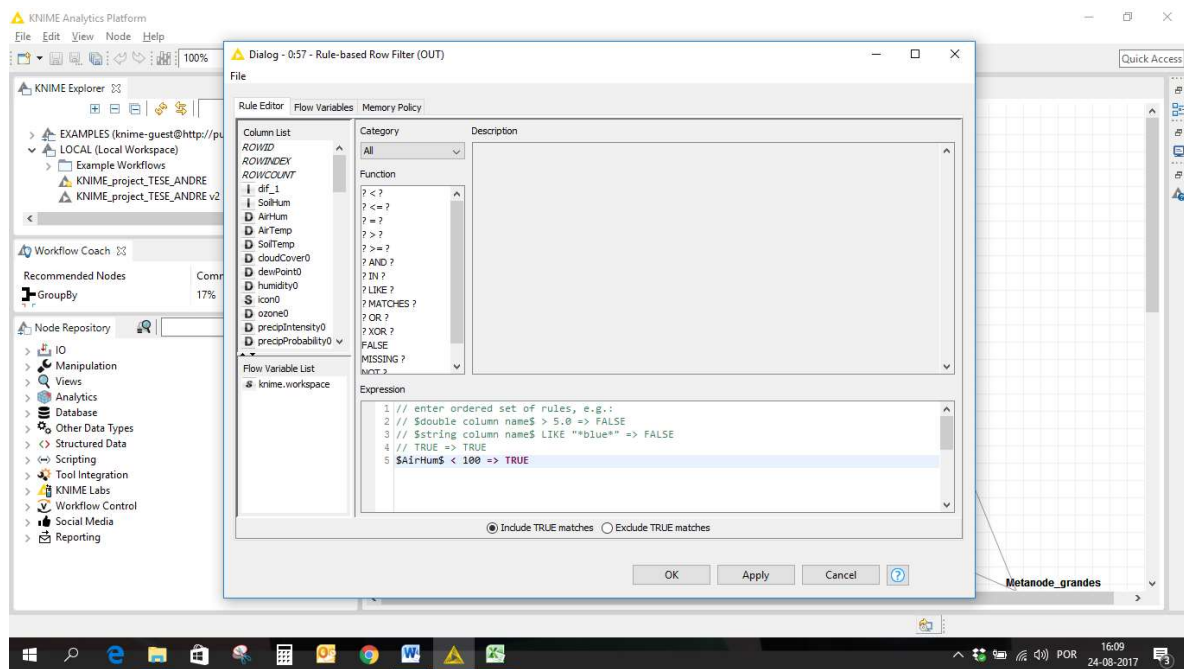
15. Guo, Z, Chi, D, Wu, J, & Zhang, W 2014, 'A new wind speed forecasting strategy based on the chaotic time series modelling technique and the Apriori algorithm', *Energy Conversion and Management*, vol. 84, pp. 140-151. Available from: 10.1016/j.enconman.2014.04.028.
16. Tao, L, & Viljanen, M 2009, 'Prediction of indoor temperature and relative humidity using neural network models: model comparison', *Neural Computing & Applications*, vol. 18, no. 4, pp. 345-357. Available from: 10.1007/s00521-008-0185-3.
17. Yang, Q, Zhu, J, Xu, X, & Lu, J 2016, 'Simultaneous control of indoor air temperature and humidity for a chilled water based air conditioning system using neural networks', *Energy & Buildings*, vol. 110, pp. 159-169. Available from: 10.1016/j.enbuild.2015.10.034.
18. Schaffer JD, Whitley D, Eshelman L (1992) Combination of genetic algorithm and neural networks: A survey of the state of art. In: *International workshop on Combinations of Genetic Algorithms and Neural Networks*, Baltimore, MD, USA, pp 1–37
19. Christian, R, & Patel, D 2015, 'Potential of k-Means Clustering-Based Fuzzy Logic for Prediction of Temperature in Ambient Atmosphere', *Arabian Journal for Science & Engineering* (Springer Science & Business Media B.V.), vol. 40, no. 1, pp. 227-234. Available from: 10.1007/s13369-014-1495-7
20. Baydaroğlu, Ö, & Koçak, K 2014, 'SVR-based prediction of evaporation combined with chaotic approach', *Journal of Hydrology*, vol. 508, pp. 356-363. Available from: 10.1016/j.jhydrol.2013.11.008.
21. Coopersmith, EJ, Minsker, BS, Wenzel, CE, & Gilmore, BJ 2014, 'Machine learning assessments of soil drying for agricultural planning', *Computers and Electronics in Agriculture*, vol. 104, pp. 93-104. Available from: 10.1016/j.compag.2014.04.004.
22. Mellit, A, Pavan, A, & Benghane, M 2013, 'Least squares support vector machine for short-term prediction of meteorological time series', *Theoretical & Applied Climatology*, vol. 111, no. 1/2, pp. 297-307. Available from: 10.1007/s00704-012-0661-7.
23. Wi-Fi Alliance® introduces low power, long range Wi-Fi HaLow™ New technology will extend Wi-Fi® solutions for the Internet of Things Las Vegas, NV, January 4, 2016
24. Gomez, C, & Paradells, J 2015, 'Urban Automation Networks: Current and Emerging Solutions for Sensed Data Collection and Actuation in Smart Cities', *Sensors* (14248220), vol. 15, no. 9, pp. 22874-22898. Available from: 10.3390/s150922874.

Anexos

Dialog - 0:2 - Column Filter (RETIRAR COLUNAS)



Anexo 1 - Variáveis retiradas



Anexo 2 - Exemplo de retirar possíveis erros/outliers

Exemplo: se a humidade superior a 100% é retirado da base de dados ou aplicar a temperatura para valores superiores a 50 graus ou inferior a -10.

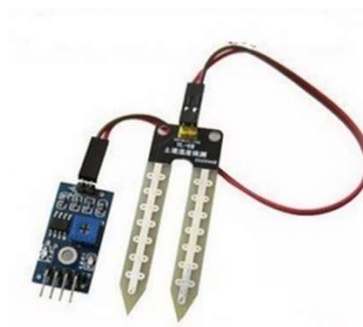


Anexo 3 - Microcontrolador utilizado - Arduino Uno Rev3

| | |
|------------------------------------|------------------------------------|
| Microcontroller | ATmega328P |
| Operating Voltage | 5V |
| Input Voltage (recommended) | 7-12V |
| Input Voltage (limit) | 6-20V |
| Digital I/O Pins | 14 (of which 6 provide PWM output) |
| PWM Digital I/O Pins | 6 |
| Analog Input Pins | 6 |
| DC Current per I/O Pin | 20 mA |
| DC Current for 3.3V Pin | 50 mA |
| Flash Memory | 32 KB (ATmega328P) |
| SRAM | 2 KB (ATmega328P) |
| EEPROM | 1 KB (ATmega328P) |
| Clock Speed | 16 MHz |
| LED_BUILTIN | 13 |
| Length | 68.6 mm |
| Width | 53.4 mm |
| Weight | 25 g |

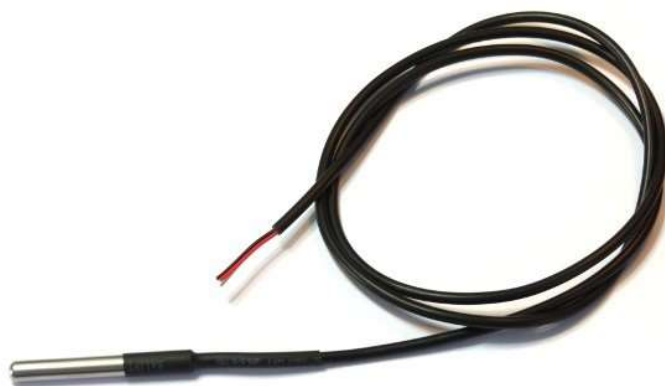
Anexo 4 - características do microcontrolador

Características do modelo - <https://store.arduino.cc/arduino-uno-rev3>



Anexo 5 - sensor de humidade do solo - G305

- Tensão de alimentação: 3.3 V ou 5 V
- Corrente de funcionamento: menos do que 20Ma
- Tensão de saída: 0-2.3 V [2.3 V é completamente imerso em água].



Anexo 6 - Sensor temperatura solo - DS18B20

- Comprimento do cabo: 100 CM
- Tamanho da bainha do aço Inoxidável: 6*50mm
- Faixa de potência de alimentação: 3.0 V a 5.5 V
- Faixa de temperatura operacional: -55 ° C a + 125 ° C (-67 ° F a + 257 ° F)
- Faixa de temperatura de armazenamento: -55 ° C a + 125 ° C (-67 ° F a + 257 ° F)
- Precisão acima da faixa de -10 ° C a + 85 ° C: ± 5 ° C



Anexo 7 - Sensor humidade e temperatura do ar - DHT22 (AM2303)

- Faixa de medição de temperatura: -40c para 80c
- Precisão de medição: 0.5 ° c
- A faixa de medição de umidade: 0-100% RH
- Precisão de medição: 2% RH




Anexo 8 - Modulo de Comunicação - NRF24L01 +

O nRF24L01 + é um transceptor de Radio Frequência de 2Mbps de alta potência e ultra baixo consumo para a banda 2.4GHz ISM (Industrial, Scientific and Medical - industrial, científica e médica).

Características:

- Voltage: 3-3.6V (recommended 3.3V)
- Maximum output power: +20dBm
- Emission mode current(peak): 115mA
- Receive Mode Current(peak): 45mA
- Power-down mode current: 4.2uA
- Sensitivity 2Mbps mode in received: -92dBm



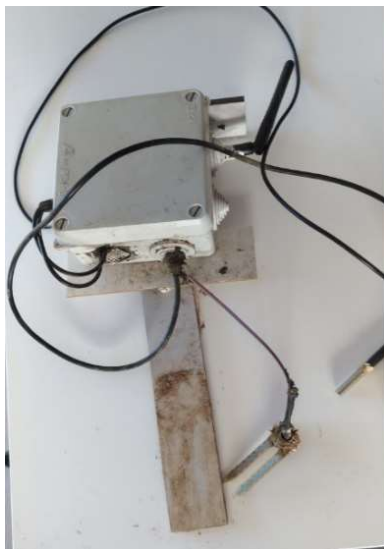
```

1  Function Convert-
2  [timezone]::Cu
3  AddSeconds($Un
4  }
5  Function ConvertT
6
7  $ports = [Sys
8  $port= new-Ob
9
10
11 do{
12     $connected =
13     try{
14         $port.open
15
16         $line = $
17         $port.clc
18         $collected
19
20         $weather
21
22         $current
23         $hour
24
25         $Current.
26         $Current.
27         $Current.
28         $Current.
29
30         ForEach (
31             $h.ti
32             $h.Te
33             $h.Ap
34             $h.de
35
36         }
37
38         $Property
39         $object
40         = New-Object -typename Pobject
41
42         $object | Add-Member -MemberType NoteProperty -Name SoilHum -Value $collected_data[1]
43         $object | Add-Member -MemberType NoteProperty -Name AirHum -Value $collected_data[2]
44         $object | Add-Member -MemberType NoteProperty -Name AirTemp -Value $collected_data[3]
45         $object | Add-Member -MemberType NoteProperty -Name SoilTemp -Value $collected_data[4]
46
47         $Properties.Name | %{
48             $object | Add-Member -MemberType NoteProperty -Name ("$_0") -Value ($Current."$_")
49         }
50
51         0..($hour.length-1) | %{
52             $i = $_
53             $Properties.Name | %{
54                 $object | Add-Member -MemberType NoteProperty -Name ("$_$_($i+1)") -Value ($hour[$i]."$_")
55             }
56         }
57
58         echo $object
59         $object | Export-CSV -delimiter ";" -.\dados.csv -append
60     }
61     catch{
62         echo $_
63         $connected = $false
64     }
65 }while($connected)

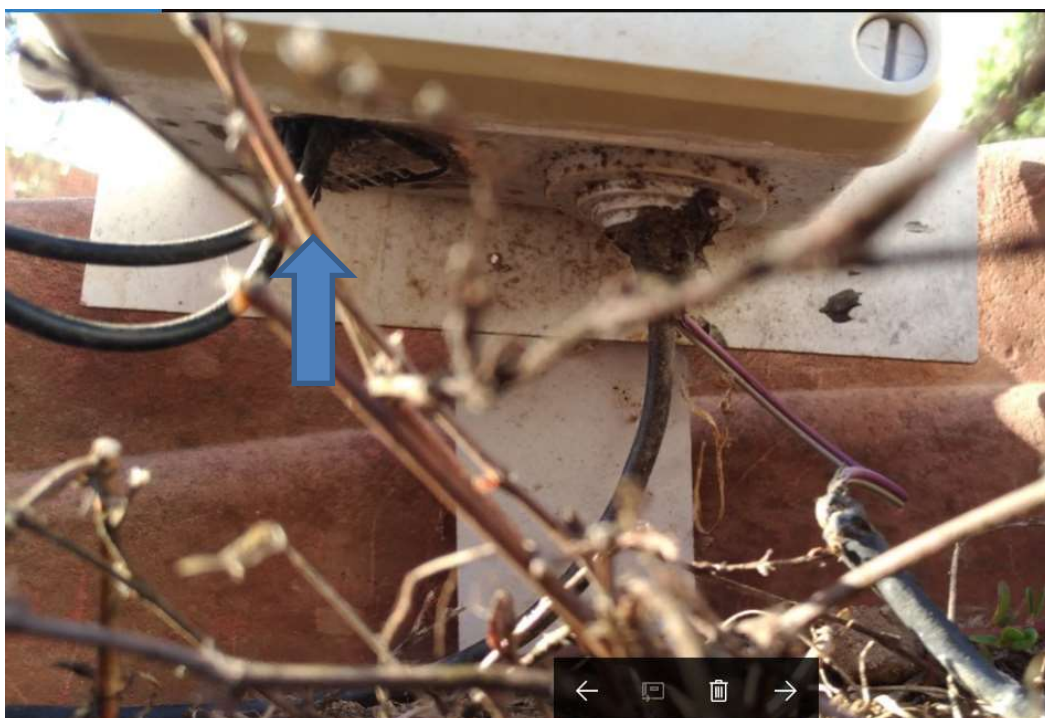
```

Anexo 9 – CÓDIGO para POWERSHELL

Anexo 10 - Nó emissor aplicado no Vaso



Anexo 11 - Nó emissor



Anexo 12 - Nó emissor - detalhe do sensor da humidade e temperatura do ar



Anexo 13 - detalhe do sensor da humidade e temperatura do solo



Anexo 14 - PC esperando por leitura de dados do receptor



Anexo 15 - PC após receber leitura do receptor - mostra dados recebidos